PCT/GB99/01061

The Patent Office

INVESTOR IN PEOPLE

09/622977

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP9 1RH

RECD 26 MAY 1999
WIPO PCT

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 18th May, 1999

Patents Form 1/77

The Patent Office

Act 1977
16)

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — A25516

2. Patent application number *(The Patent Office will fill in this part)* — 9807745.6

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

BRITISH TELECOMMUNICATIONS public limited company
81 NEWGATE STREET
LONDON, EC1A 7AJ, England
Registered in England: 1800000

Patents ADP number *(if you know it)* 1867002

If the applicant is a corporate body, give the country/state of its incorporation — UNITED KINGDOM

4. Title of the invention — TELECONFERENCING SYSTEM

5. Name of your agent *(if you have one)* — Timothy Guy Edwin LIDBETTER

"Address for Service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

BT GROUP LEGAL SERVICES
INTELLECTUAL PROPERTY DEPARTMENT
HOLBORN CENTRE
120 HOLBORN
LONDON, EC1N 2TE

Patents ADP number *(if you know it)* 1867001 0699c+9300/

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day/month/year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.

See note (d))

YES

following items you are filing with this form.
Do not count copies of the same document

Continuation sheets of this form

Description 9 /

Claim(s) 2 / A.

Abstract 1 /

Drawing(s) 11 + 11 /

10. If you are also filing any of the following, state how may against each item

Priority Documents

Translations of priority documents

Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*

Request for preliminary examination and search *(Patents Form 9/77)*

Request for substantive examination *(Patents Form 10/77)*

Any other documents *(please specify)*

11. I/We request the grant of a patent on the basis of this application.
Signature(s) — Date:

08 APRIL 1998

Timothy Guy Edwin LIDBETTER, Authorised Signatory

12. Name and daytime telephone number of person to contact in the United Kingdom

Bhavna Vasani 0171 492 8146

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

TELECONFERENCING SYSTEM

This invention relates to audio teleconferencing systems. These are systems in which three or more participants, each having a telephone connection, can participate in a multi-way discussion. The essential part of a teleconference system is called the conference "bridge", and is where the audio signals from all the participants are combined. Conference bridges presently function by receiving audio from each of the participants, appropriately mixing the audio signals, and then distributing the mixed signal to each of the participants. All signal processing is concentrated in the bridge, and the result is monaural (that is, there is a single sound channel). This arrangement is shown in Figure 1, which will be described in detail later. The principal drawback with this system is that the audio quality is monophonic, generally poor, it is very difficult to determine which participants are speaking at any one time, especially when the number of participants is large.

According to the invention, there is provided a teleconferencing system comprising a conference bridge having a multichannel connection to each customer equipment, the customer equipment having means to separately process each channel to provide an output, preferably spatialised, representing each of the other participants. Preferably the conference bridge comprises a concentrator, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels. This reduces the capacity required by the multichannel connection. The control information identifying the active channels may be carried in a separate control channel, or as an overhead on the active subset of channels. In a preferred arrangement the channel representing a given participant is excluded from the output provided to that participant. This may be achieved by excluding that channel from the processing in the customer equipment, but is preferably achieved by excluding it from the multichannel transmission from the bridge to that participant, thereby reducing further the capacity required by the multichannel connection.

Exemplary embodiments of the invention will now be described, by way of example, with reference to the drawings, in which:

Figure 1 illustrates a conventional teleconference system;

Figure 2 illustrates a spatial audio teleconference system according to one embodiment of the invention;

Figure 3 illustrates a N-channel speech decoder used in the embodiment of Figure 2;

Figure 4 illustrates a N-Channel audio spatialiser used in the embodiment of Figure 2;

Figure 5 illustrates a second embodiment of the invention;

Figure 6 illustrates how the invention may be used with conventional PSTN channels;

Figure 7 illustrates a variant of the invention for use with a video conference system;

Figure 8 illustrates a voice switched concentrator which may be used in the embodiments of the invention.

Figures 9, 10, and 11 illustrate various echo cancellation techniques.

In the conventional system illustrated in Figure 1 the conference bridge located in the exchange equipment 100 receives signals from the various customer equipments 10, (20, 30 not shown) in response to sounds detected by respective microphones 11, 21, 31 etc. These signals are transmitted over the telephone network (1), to the exchange 100 at which the bridge is established. Generally the signals will travel by way of a local exchange (not shown) in which the analogue signals are converted to digital form, usually employing linear companding such as "A law" (as used for example in Europe) or "mu-Law" (as used for example in the United States of America) for onward transmission to the bridge exchange 100. On arrival at the bridge exchange 100, the bridge passes each incoming signal 11, 21, 31 through a respective digital converter 111, 112, 113 to convert them from A Law to linear digital signals, and then passes the linear signals to a digital combiner 120 to generate a combined signal. This combined signal is re-converted to A law in a further digital converter 110, and the resulting signal transmitted over the telephone network (2) to each customer equipment 10, (20, 30) for conversion to sound in respective loudspeakers 12, 22, 32 etc. In this way the exchange equipment 100 acts as a "bridge" to allow one or more customer equipments 32 to connect into a simple two-way connection between customer equipments 10, 20.

The systems illustrated in Figures 2 to 8 replace the conventional conference bridge system of Figure 1 with a multicast system in which several channels can be transmitted to each participant, using a multi-channel link comprising an uplink 3, and a downlink comprising a control channel 4 and a digital audio downlink 5 comprising several channels 51, 52. Participants with suitable equipment can then process these channels 51, 52 in various ways as will be described.

The transmission medium used for the uplink 3 and downlink 4,5 can be any suitable medium. ISDN (Integrated Services Data Network) technology or LAN (Local Area Network) - respectively public and private data networks - are the favoured transmission options since they provide adequate data rate and low latency - delays due to coding and transmission buffering. However, they are expensive and so far have a low penetration in the market place. Internet Protocol techniques are more widely used, but have poor latency and unreliable data rates. Being packet systems, they are less suited to voice applications. It is also possible to use the conventional PSTN (public switch telephone system) with a speech band modem. The latest internet type modems provide up to 56kbit/s downstream (links 4,5: digital network down to the customer via local loop), and up to 28.8kbit/s upstream (link 3). They are low cost and are commonly bundled into PC packages. Ideally a system should be able to work with all of the above, and with standard analogue PSTN available as a backup.

The signal mixing can take place either in the user's terminal equipment, or in a centralised processing platform as shown in Figure 2. In Figure 2 the customer equipment 10 contains a microphone 11 and loudspeaker system 12 as before. However, the loudspeaker system 12 is a spatialised system - that is, it has two or more channels to allow sounds to appear to emanate from different directions. This may take the form of stereophonic headphones, or a more complex system such as disclosed in United States Patents 5533129 (Gefvert), 5307415 (Fosgate), article *"Spatial Sound for Telepresence"* by M.Hollier, D. Burraston, and A. Rimell in the *British Telecom Technology Journal,* October 1997 or the applicant's own pending European Patent Application 97304218.7 filed on 17th June 1997.

The output from the microphone 11 is encoded by an encoder 13 forming part of the customer equipment 10, and transmitted over the uplink 3 to the

exchange equipment 100. Here it is combined with the other input channels 21, 31 from the other participants into a concentrator 230 which combines the various inputs into an audio signal having a smaller number of channels 51, 52. These channels are transmitted over multiple-channel digital audio links 5 to the customer equipments 10, (20, 30) where they are first decoded by respective decoders 14, 24, 34 and provided to a spatialiser 15 for controlling the mixing of the channels to the generate a spatialised signal in the spaeker equipment 12.

The concentrator 230 selects from the input channels 11, 21, 31 those carrying useful information - typically those carrying speech, and passes only these over the return link 5. This reduces the amount of information to be carried. A control channel 4 carries data identifying which channels were selected. The spatialiser 15 uses data from the control channel to identify which of the original sound channels 11, 21, 31 it is receiving, and on which of the "N" channels 51, 52 in the audio link each original channel is present, and constructs a spatialised signal using that information. The spatialised signal can be tailored to the individual customer, for example the number of talkers in the spatialised system, the customer's preferences as to where in the spatialised system each participant is to appear to be located, and which channels to include; for example the original talker or a simultaneous translation. In particular, the user may exclude the channel representing his own input 11.

Transmission efficiency is achieved because only the active subset N of the total number of channels M are transmitted at any one time. The subset is chosen using a voice controlled dynamic channel allocation algorithm in the N:M concentrator 230. A possible implementation of this is shown in Figure 8. Each input channel 11,21,31 is monitored by a respective analyser 231, 232, 233. As shown for analyser 231, the signal is subjected to a speech detection and analysis process 231b. This detects whether speech is present on the respective input 11, and gives a confidence value, indicative of how likely the signal contains speech. This ensures that low-level background speech is given a lower weight than speech clearly addressed to the microphones 11, 21, 31 etc. A value is also given for level, to ensure speech directed to the microphone is preferred over background noise, and the level information can be passed to the spatialistion system to select a coding algorithm appropriate to the information in the speech. In order to detect

and process the speech in the signals they first need to be decoded in a decoder 231a (this may be dispensed with if the speech detection system 231b can operate with digitally encoded signals).

A voting algorithm 234 then selects which of the inputs 11, 21, 31 have the clearest speech signals and controls a switch to direct each of those input channels 11, 21, 31 which have been selected to a respective one of the output channels 51, 52. Similar algorithms are used in Digital Circuit Multiplication Equipment (DCME) systems in international telephony circuits. Data relating the audio channels' content to the conference participants, and therefore the correspondence between the input channels 11, 21, 31 and output channels 51, 52 is transmitted over the control channel 4. Alternatively, this data can be embedded in the encoded audio data.

When there are fewer talkers identified than there are available output channels 51, 52, signal quality can be improved by using a less compressed digitisation scheme for those input channels selected, thereby using more than one output channel 51, 52 for each input channel selected. Telephone quality speech may be achieved at 8kbits/s, allowing eight talkers to be accommodated if the system has a 64kbit/sec capability. Should fewer talkers be detected, the 64kbit/s capability may be used instead to provide four 16 kbit/s audio channels, capable of carrying 'good' quality speech, or a mixture of channels at different bit rates, to allow the coding rates to be selected according to the initial signal quality, or so that the main talker could be passed at higher quality than the other talkers. Layered coding schemes can be used to allow graceful switching between data rates.

The N-channel de-multiplexer and speech decoder 14 is shown in Figure 3. This receives the channels 51, 52, 53 etc carried in the audio downlink 5 and separates them in a demultiplexer 140. Each channel 51, 52, etc is then separately decoded in a respective decoder 141, 142, 143, etc for processing by the spatialiser 15. The decoders 141, 142, etc may operate according to different processes according to the individual coding algorithms used, under the control of the control signals carried in the control channel 4.

A composite signal consisting of the summation of all input signals could also be transmitted. Such a signal could be used by users having monaural

receiving equipment, and may also be used by the spatialiser to generate an ambient background signal. Alternatively the spatialiser 15 may replace any of the channels 11, 21, 31 not selected by the concentrator 230, and therefore not represented in the N channel link 5, by "comfort noise"; that is low-level white noise, to avoid the aural impression of a void which would be occasioned by a complete absence of signal.

The customer equipment 10 can be implemented using a desktop PC. Readily available PC sound card technology can provide the audio interface and processing needed for the simpler spatialising schemes. The more advanced sound cards with built in DSP technology could be used for more advanced schemes. The spatialiser 15 can use any one of a number of established techniques for creating an artificial audio environment as detailed in the Hollier et al article referred to above. Spatialisation techniques may be summarised as follows. The simplest technique is "panning", where each signal is replayed with appropriate weighting via two or more loudspeakers such that it is perceived as emanating from the required direction. This is easy to implement, robust and may also be used with headphones.

"Ambisonic" systems are more complicated and employ a technique known as wavefront reconstruction to provide a realistic spatial audio percept. They can create very good spatial sound, but only for a very small listening area and are thus only appropriate for single listeners. For headphone listening, "binaural" techniques can be used to provide very good spatialisaton. These use head-related transfer function (HRTF) filter pairs to recreate the soundfield that would have been present at the entrance to the ear canal for a sound emanating from any position in 3D space. This can give very good spatialisation and may be extended for use with loudspeakers, when it is known as "transaural". As with ambisonic systems, the correct listening position is very small. Any of these spatialisation techniques may be used with the present invention.

The output of several spatialisers may be combined as shown in Figure 4, which shows a spatialiser group for a stereophonic output having left and right channels 12L, 12R. Each channel 51, 52, 53 is fed to a respective spatialiser 151, 152, 153 which, under the control of a coefficient selector 150 control by the signals in the control channel 4, transmits an output 151L, 151R etc to each of a

series of combiners 15L, 15R. The processing used to create the outputs 151L, 151R etc is operated under the control of the signal 4 such that each channel appears as a virtual sound source, having its own location in the space around the listener.

The positions of virtual sources in three dimensional space could be determined automatically, or by manual control, with the user selecting the preferred positioning for each virtual sound source. For a video conference the positioning can be set to correspond with the appropriate video picture window. The video images may be sent by other means, or may be static images retrieved from a local storage by the individual user.

If the spatialised sound is relayed via loudspeakers 12, rather than headphones, it will be necessary to prevent signals from the loudspeakers 12 being picked up by the microphone 11, re-transmitted and being heard as an echo at the distant sites 20, 30 etc. A technique for achieving this will be described later, with reference to Figure 11.

Figure 5 shows an alternative arrangement to that of Figure 4, in which the spatialisation is computed in the conference 'bridge'. Each conference participant gets the same spatialised signals, thus simplifying the customer equipment. Figure 5 is similar in general arrangement to Figure 2, except that the decoder 14 and spatialiser 15 are part of the exchange equipment 200. The output from the spatialiser 15 is passed to an encoder 18 which transmits the required number of audio channels (e.g. two for a stereo system) to each customer 10, 20, 30. This requires the number of channels in the downlink 5 to be equal to the number of audio channels in the spatialisation systems' outputs, instead of the number selected by the concentrator (plus the control channel 4) as in the embodiment of Figure 2. It also simplifies the customer equipment 10. However, this arrangement requires all customer installations 10, 20, 30 to have similar spatialisation systems, and in particular the same number of audio channels. It would also be more difficult to remove a talker's own voice from the signal he receives. Echo control would also be more complicated, and channel coding may degrade the spatialisation.

Conventional analogue connections could be included in the conference by providing each analogue connection 43, 45 to the 'bridge' 200 with an encoder

42, as shown in Figure 6, to provide an input 41 to the concentrator. The output 5 of the concentrator 230 is also decoded and combined in a unit 44 to provide a monaural conference signal 45 to the analogue user 40.

In the embodiments of Figures 2 to 4 and Figure 5, if loudspeakers are used there is a need to control acoustic feedback ("echo") between the loudspeaker 12 and the microphone 11, which will result in signals being retransmitted back into the system. This will result in each user hearing one or more delayed versions of each signal (including their own transmissions) arriving from the other users. For a monophonic system echo control is usually done using an echo canceller as shown in Figure 9. The echo signal, represented by D is caused by the acoustic path J between the loudspeaker 12 and microphone 11 of equipment 10 in room B. The cancellation is achieved in an echo control unit 16 by using an adaptive filter to create a synthetic model of the signal path such that the echo may be removed by subtraction. The signal E, returned to equipment 20 in room A, is now free of echoes, containing only sounds that originated in Room B. The optimum modelling of the acoustic path J is usually achieved by the adaptive filter in a manner such that some appropriate function of the signal E is driven towards zero. Echo control using adaptive filters in this manner is well known.

Multi-channel echo cancellation, as shown in Figure 10 for two channels, is more complex since there are two input channels 51, 52 and therefore two loudspeakers 12L, 12R. It is therefore necessary to model two echo paths K and L for each of the two return channels 3L, 3R. (The process is only shown for return channel 3L, using microphone 11L). Correct echo cancellation is only achieved if adaptive filters 161L, 162L model the signal paths K and L respectively. (Two further filters 161R, 162R are required for the other return channel 3R) However, it is not possible to find a correct model for each path K, L independently without some difficult and expensive signal processing as described in *"A better understanding and an improved solution to the specific problems of stereophonic echo cancellation"* (IEEE Transactions on speech and Audio processing, Vol 6, no 2 March 1998. Authors: J Benesty, D R Morgan and M M Sondhi).

The system described above with reference to Figure 4 employs linear artificial spatialisation techniques. Figure 11 shows how this, and the fact that the

echo from each loudspeaker 12L, 12R combines linearly at each microphone 11L, (11R, not shown), allows echo cancellation to be provided for each output channel 3L, (3R) by having a separate adaptive filter 161L, 162L, 163L, (161R, 162R, 163R) on each input channel 51, 52, 53. Thus the adaptive filter 161L will model the combination of the spatialiser 151 for the channel 51, and the echo path between the loudspeakers 12L and 12R and the microphone 11L.

The invention could be applied to a conference situation in which there are several participants at each location, such as the video conference shown in Figure 7. Close microphones 11a, 11b, 11c, for example of the "tie-clip" type, are used to pick up the sound from each individual talker, and a talker location system 60 is used to keep track of their spatial position. The talker location system 60 may comprise a system of microphones which can identify the positions of sound sources. Relating the position of a sound source to that of the tie clip microphone 11 currently in use makes it possible to learn the position of each talker by audio means alone. Alternatively, the system may detect the position of each user by means such as optical recognition of a badge carried by each user. In either case, the position data is passed to the far end (Room B), where correct spatialisation is reconstructed, for output by loudspeakers 12L, 12M, 12R etc. This would achieve a true spatial conference and overcome the associated echo control problems, since the "tie clip" microphones 31a, 31b, 31c have a limited range and will not detect the outputs from the loudspeakers 32 in the same room.

CLAIMS

1. A teleconferencing system comprising a conference bridge having a multichannel connection to each customer equipment, at least one customer equipment having means to separately process each channel to provide a plurality of outputs, each output representing one of the other participants.

2. A system according to claim 1, wherein the customer equipment has means to combine the outputs representing each participant to provide a spatialised output in which each participant is represented by a virtual sound source.

3. A system according to claim 1 or 2, wherein, the conference bridge comprises a concentrator, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

4. A system according to any preceding claim, wherein the channel representing a given participant is excluded from the output provided to that participant.

5. A system according to claim 4, comprising means in the customer equipment for excluding the said channel from the processing.

6. A system according to claim 4, comprising means for excluding the said channel from the multichannel transmission from the bridge to the respective participant.

7. A system according to any preceding claim, the customer equipment having echo cancellation means comprising means for detecting correlations between the output signal from the customer equipment and the signals carried on individual input channels to the customer equipment representative of other users,

such correlations being indicative of acoustic feedback at the customer equipment, and means for cancelling such feedback signals in the output signal.

8. A system according to claim 7, wherein the customer equipment comprises, for each channel of the output signal, a plurality of adaptive filters, each adaptive filter being arranged to model the echo path between a respective input channel and the respective output channel, and
for each output channel there being provided a combiner for adding the outputs of the respective plurality of adaptive filters to generate an echo cancellation signal for the respective output channel.

9. A method of providing teleconferencing services to a plurality of customer equipments, in which a multichannel connection is provided from a conference bridge to each customer equipment, in which at least one customer equipment processes each channel separately to provide a plurality of outputs, each representing one of the other participants.

10. A method according to claim 9, wherein the conference bridge identifies the currently active input channels and transmits only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

ABSTRACT

TELECONFERENCING SYSTEM

A teleconferencing system comprises a conference bridge 100 having a multichannel connection 5 to each customer equipment 10. The customer equipment 10 has means 14 to separately process each channel 51, 52, 53 to provide one output representing each of the other participants. These outputs can be combined in a spatialiser 15 to provide a spatialised output 12 in which each participant is represented by a virtual sound source. The conference bridge 100 comprises a concentrator 230, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

Figure (2)

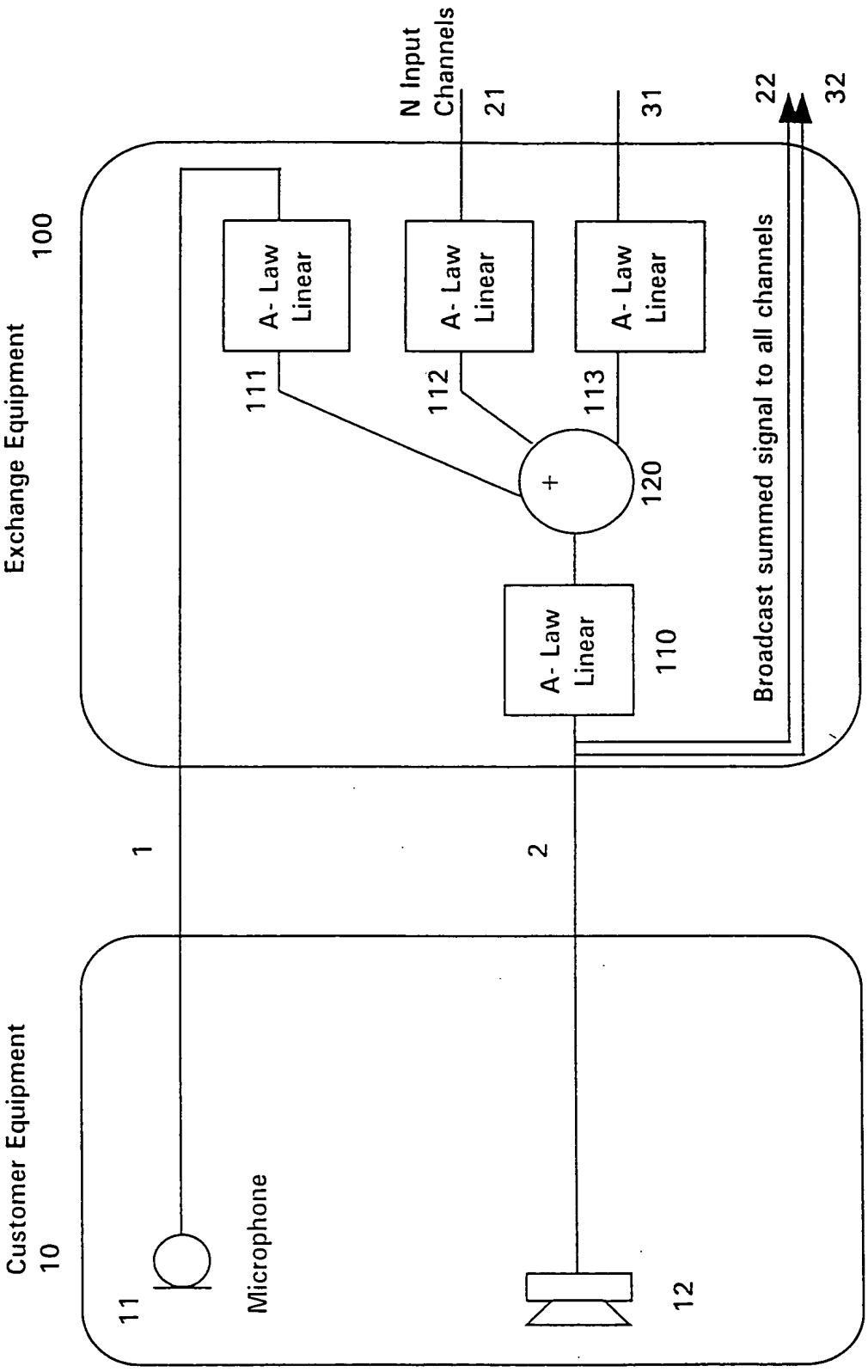
Customer Equipment
10
Microphone
11
12
1
2
Exchange Equipment
100
A- Law Linear
111
A- Law Linear
112
A- Law Linear
113
+
120
A- Law Linear
110
N Input Channels
21
31
22
32
Broadcast summed signal to all channels

Figure 1

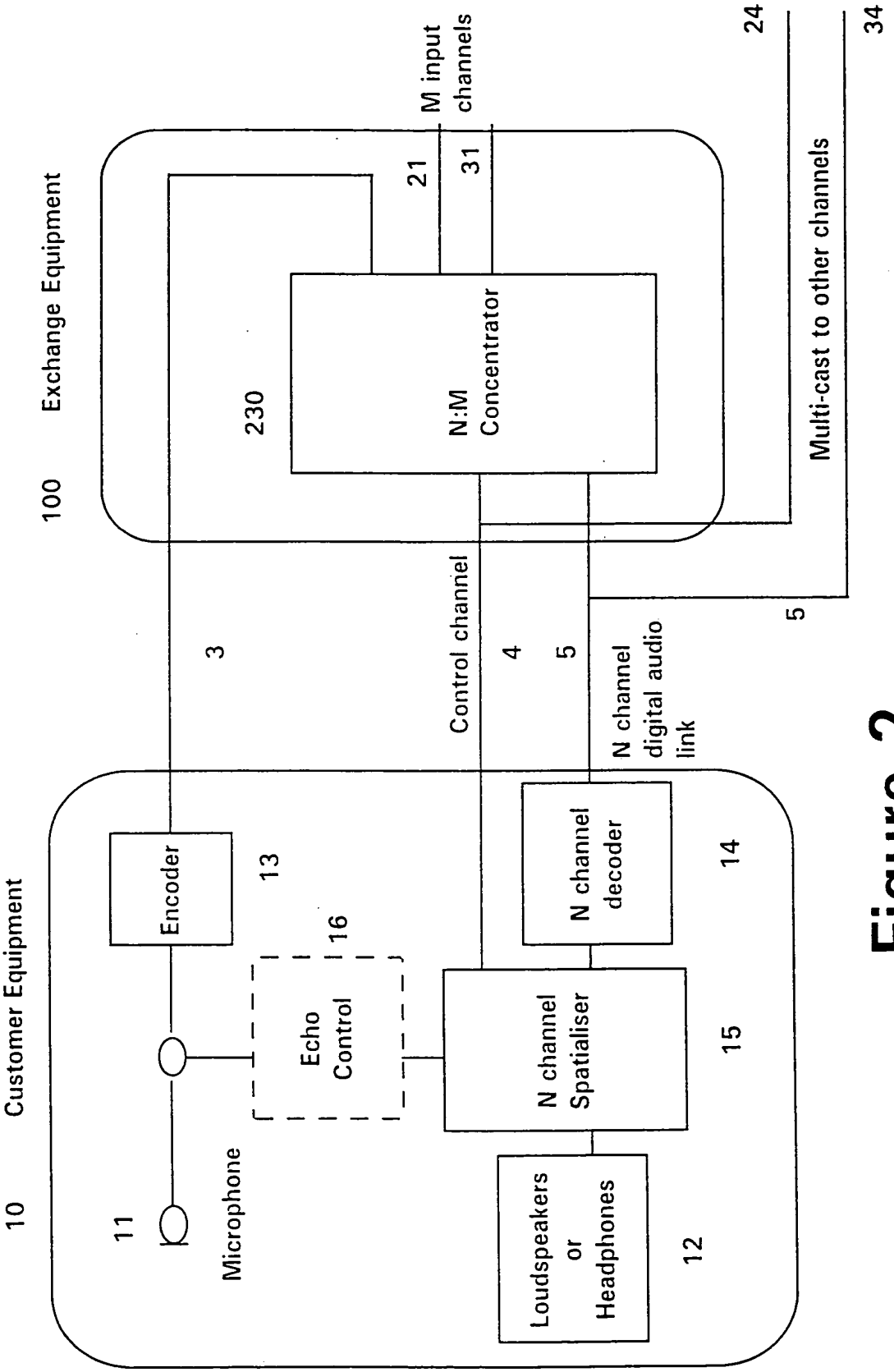
Customer Equipment
10
Microphone
11
Encoder
13
Echo Control
16
Loudspeakers or Headphones
12
N channel Spatialiser
15
N channel decoder
14
3
Control channel
4
5
N channel digital audio link
Exchange Equipment
100
230
N:M Concentrator
21
31
M input channels
5
Multi-cast to other channels
24
34

Figure 2

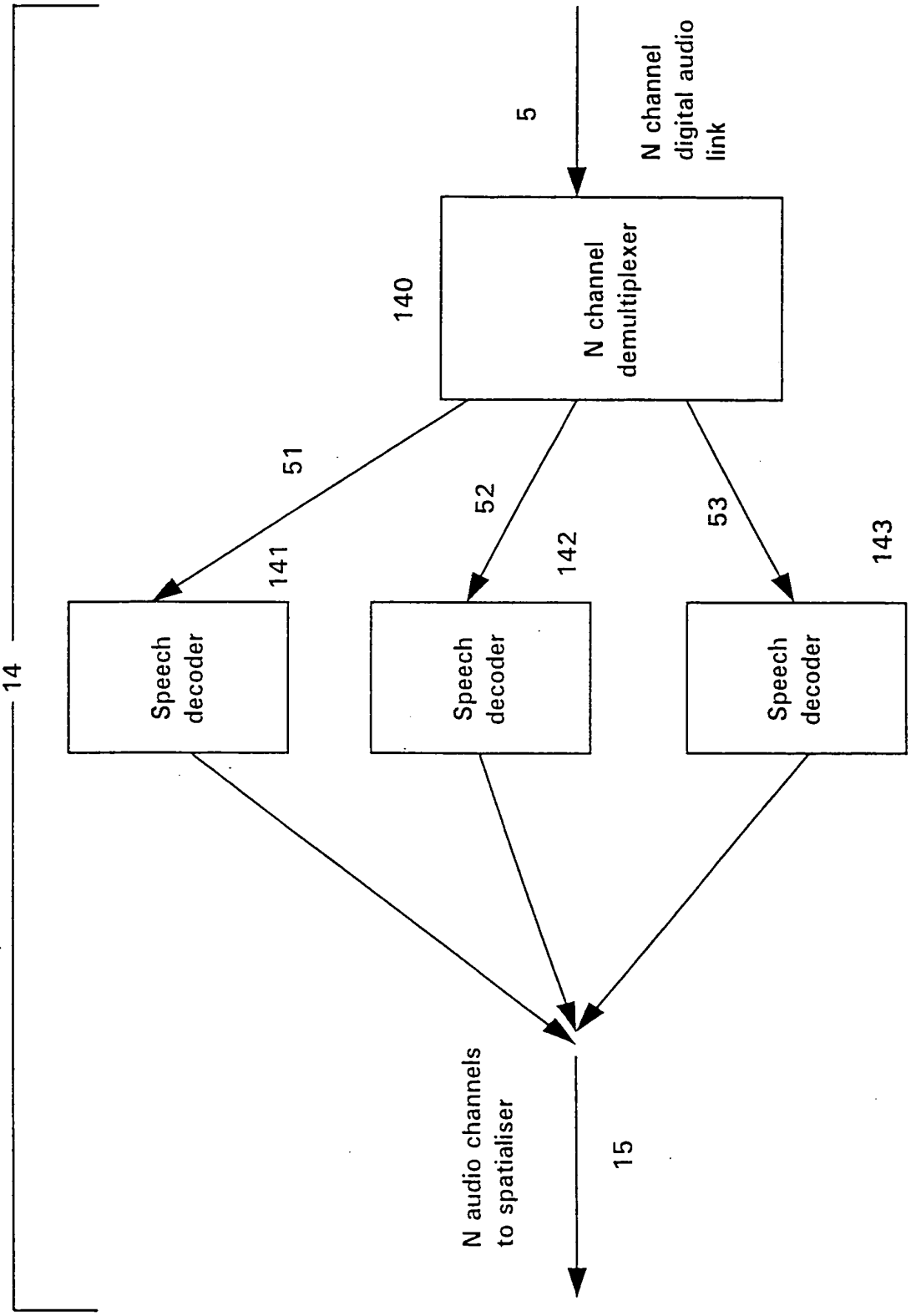
14
5
N channel digital audio link
140
N channel demultiplexer
51
52
53
141
142
143
Speech decoder
Speech decoder
Speech decoder
N audio channels to spatialiser
15

Figure 3

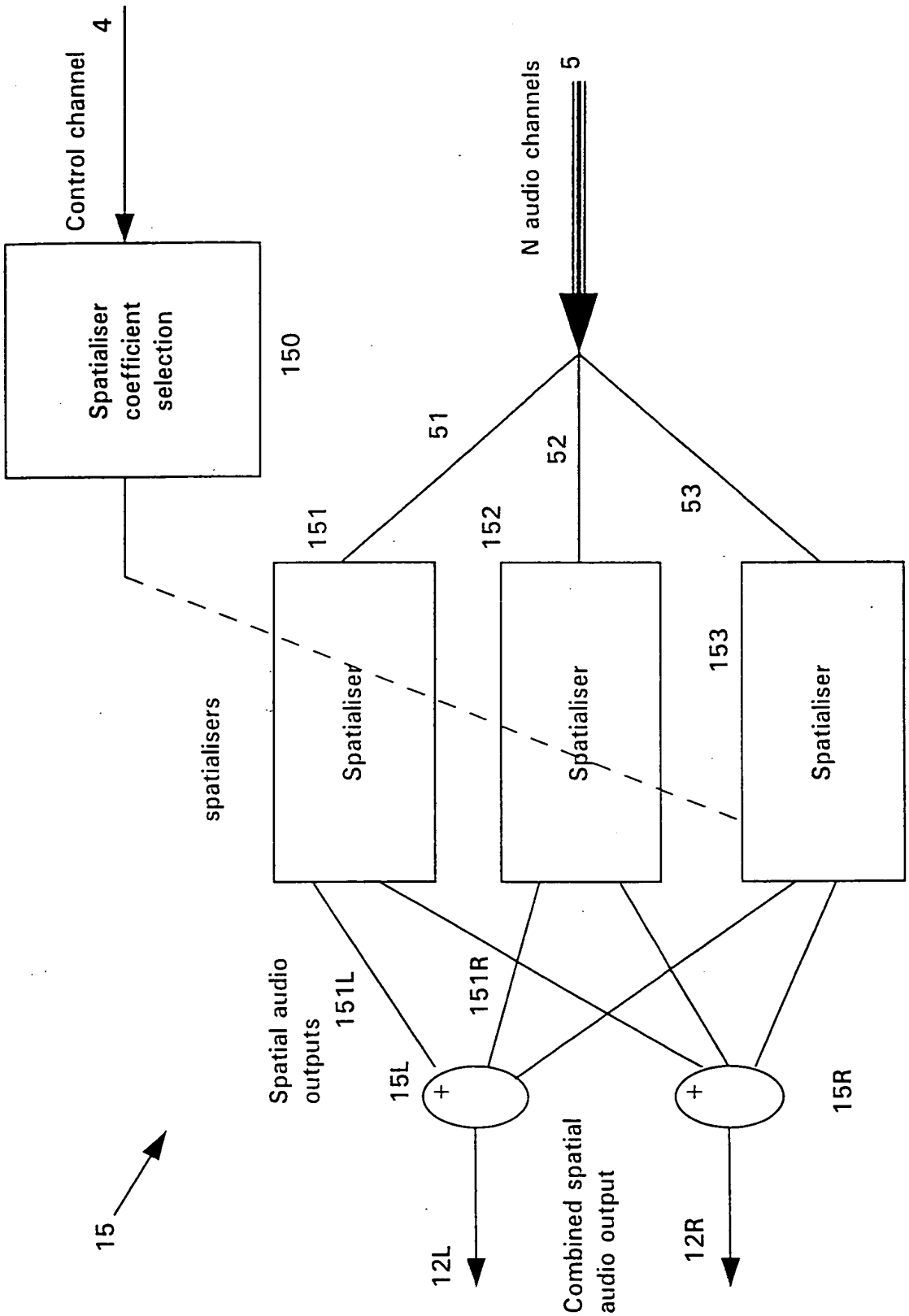
Control channel
4
Spatialiser coefficient selection
150
N audio channels
5
51
52
53
151
152
153
spatialisers
Spatialiser
Spatialiser
Spatialiser
Spatial audio outputs
151L
151R
15L
15R
12L
12R
Combined spatial audio output
15

Figure 4

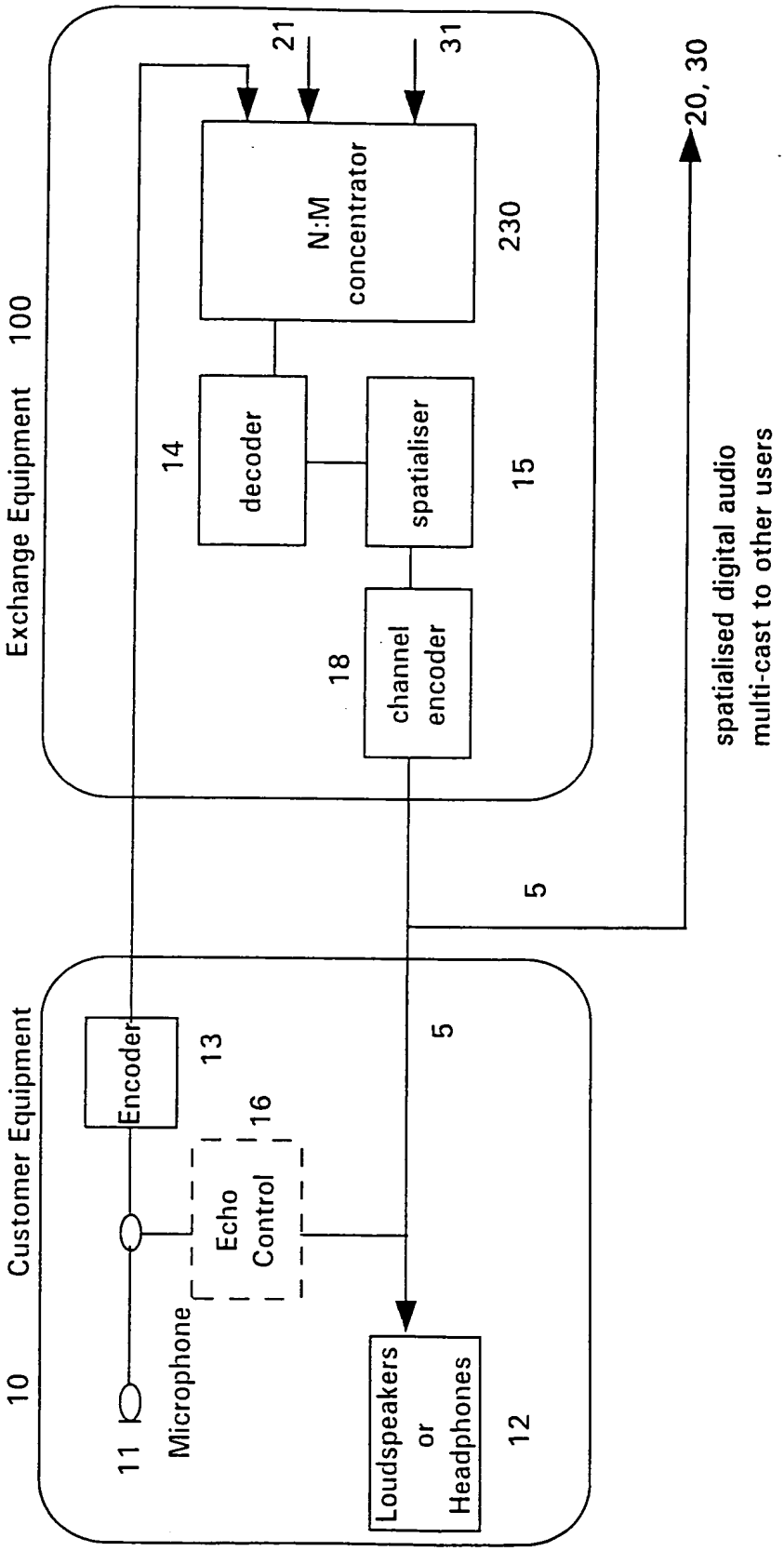
Exchange Equipment 100
21
31
N:M concentrator
230
decoder
14
spatialiser
15
channel encoder
18
20, 30
spatialised digital audio multi-cast to other users
5
Customer Equipment
10
Encoder
13
Echo Control
16
Microphone
11
Loudspeakers or Headphones
12
5

Figure 5

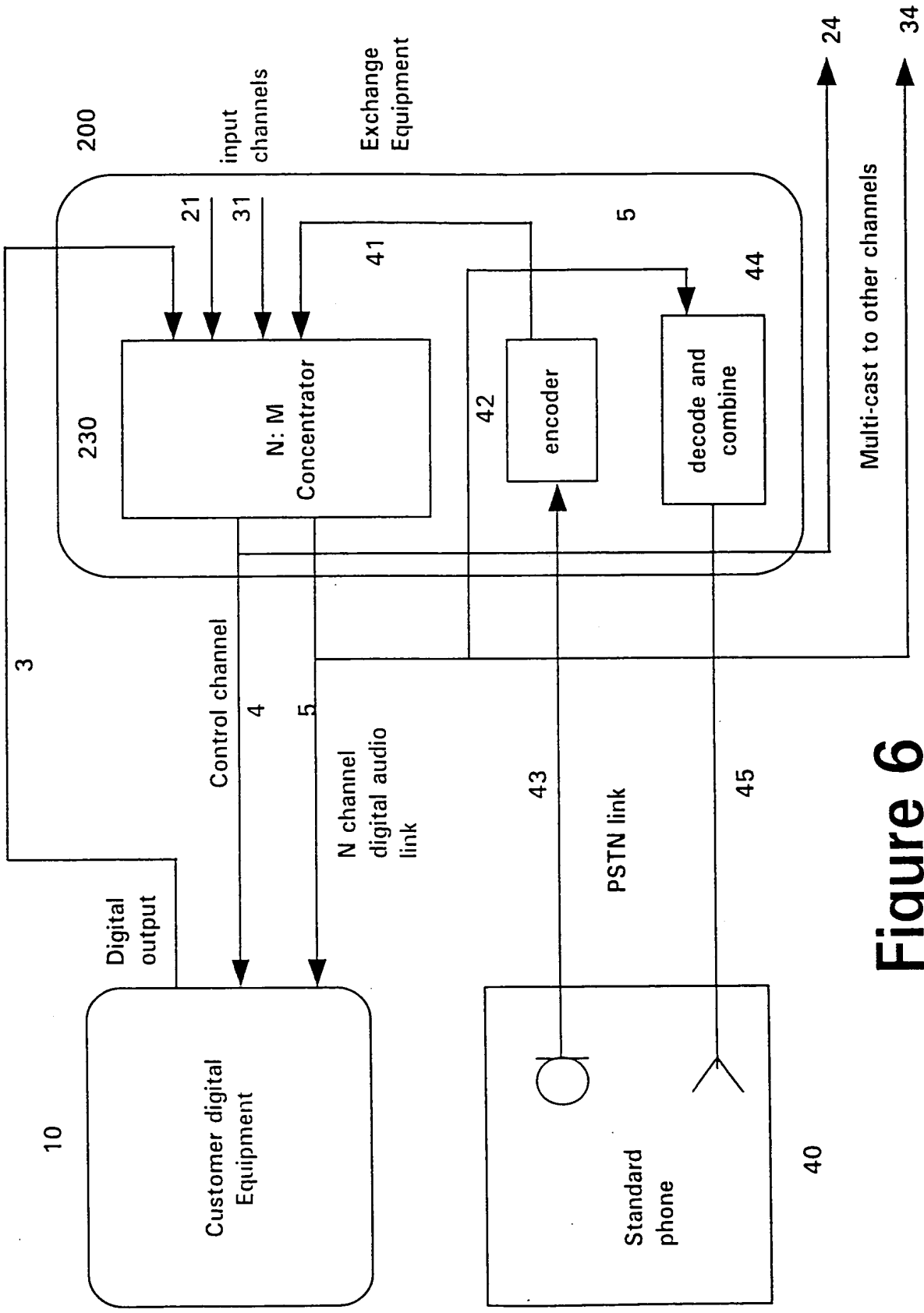
200
input channels
Exchange Equipment
24
34
21
31
41
5
44
230
N: M Concentrator
42
encoder
decode and combine
Multi-cast to other channels
3
Control channel
4
5
N channel digital audio link
43
PSTN link
45
Digital output
10
Customer digital Equipment
Standard phone
40

Figure 6

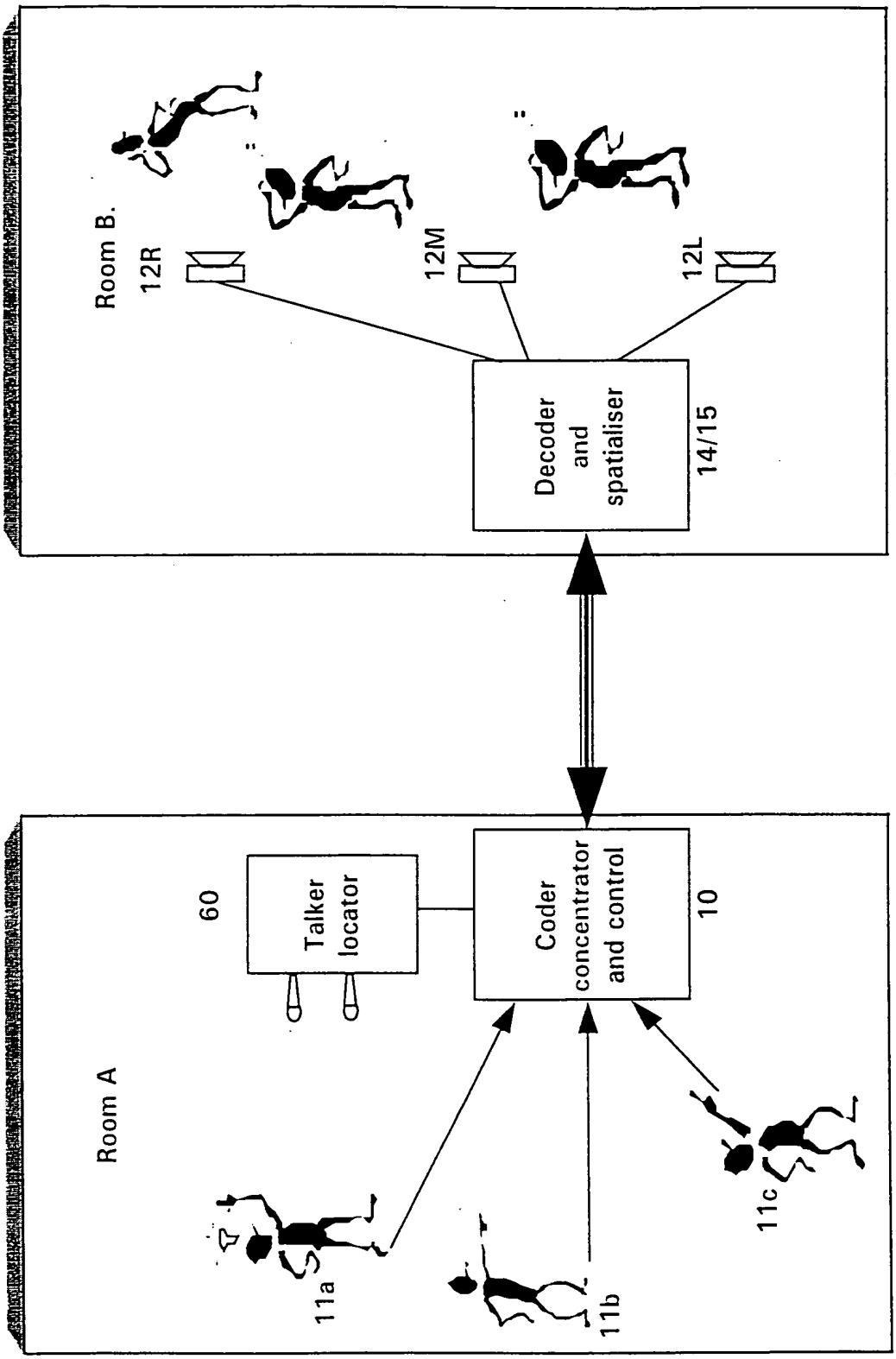
Room B.
12R
12M
12L
Decoder and spatialiser
14/15
Room A
60
Talker locator
Coder concentrator and control
10
11a
11b
11c

Figure 7

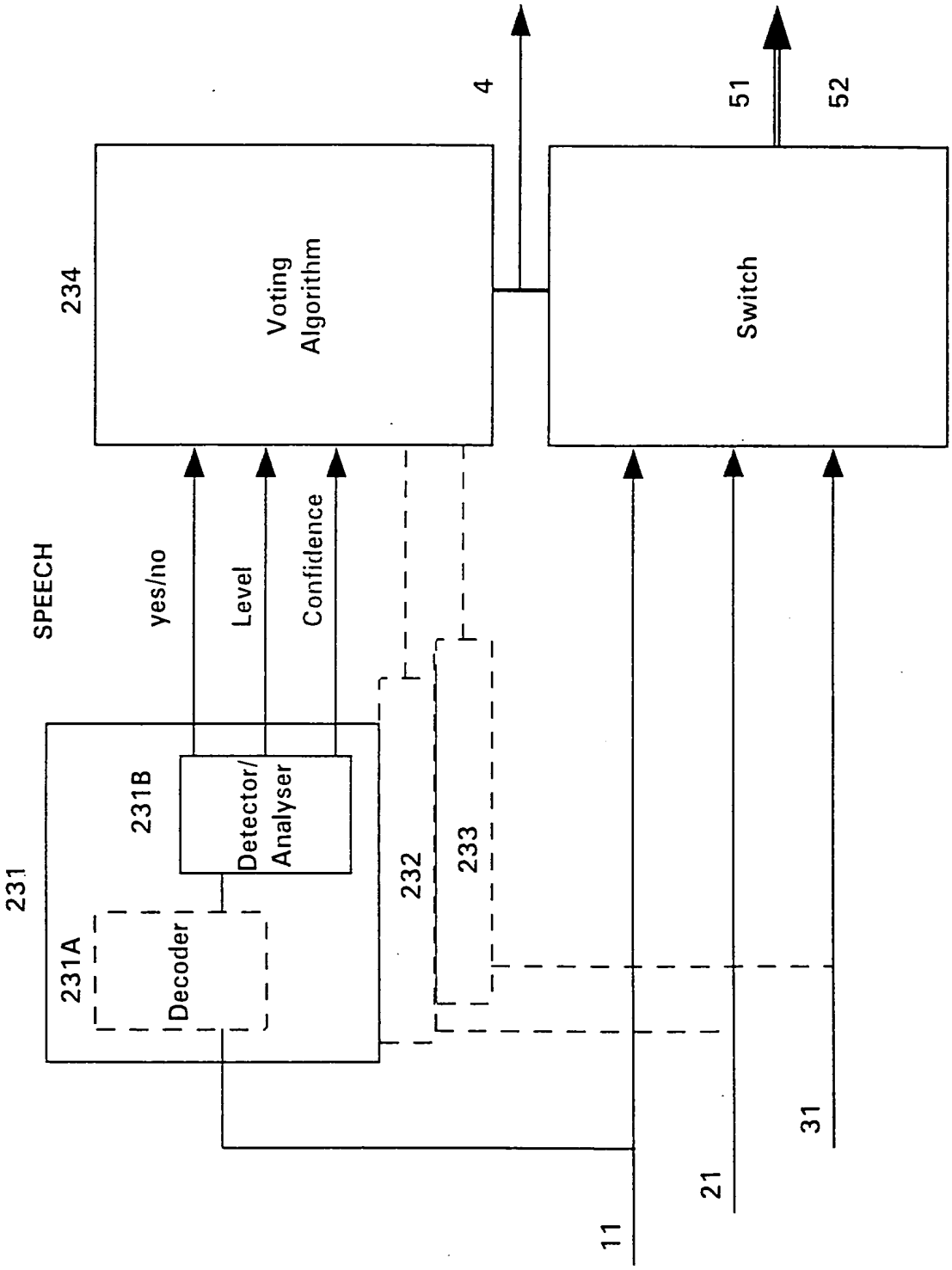
SPEECH
yes/no
Level
Confidence
231
231A
Decoder
231B
Detector/
Analyser
232
233
234
Voting
Algorithm
4
51
52
Switch
235
11
21
31

Figure 8

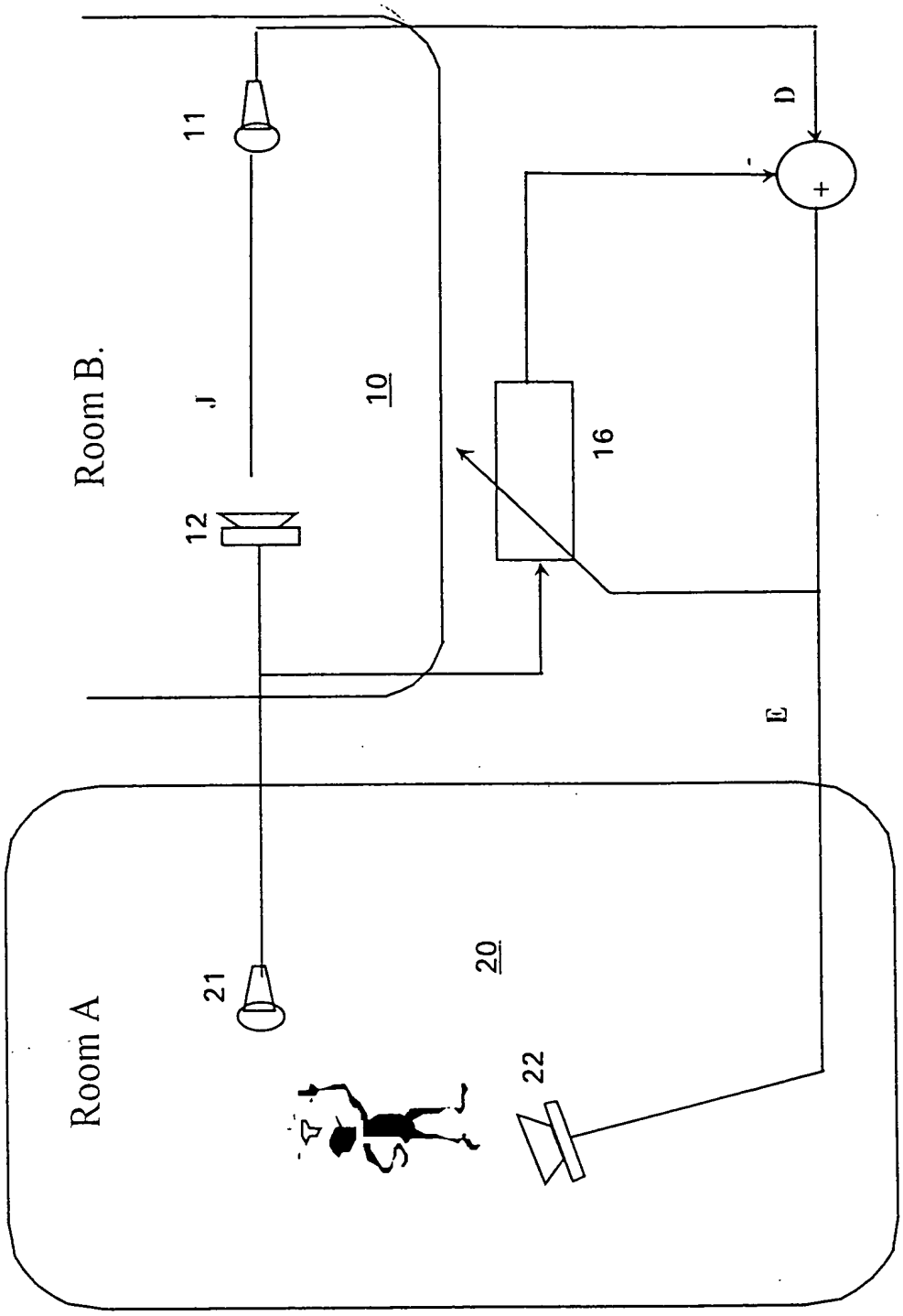

Figure 9

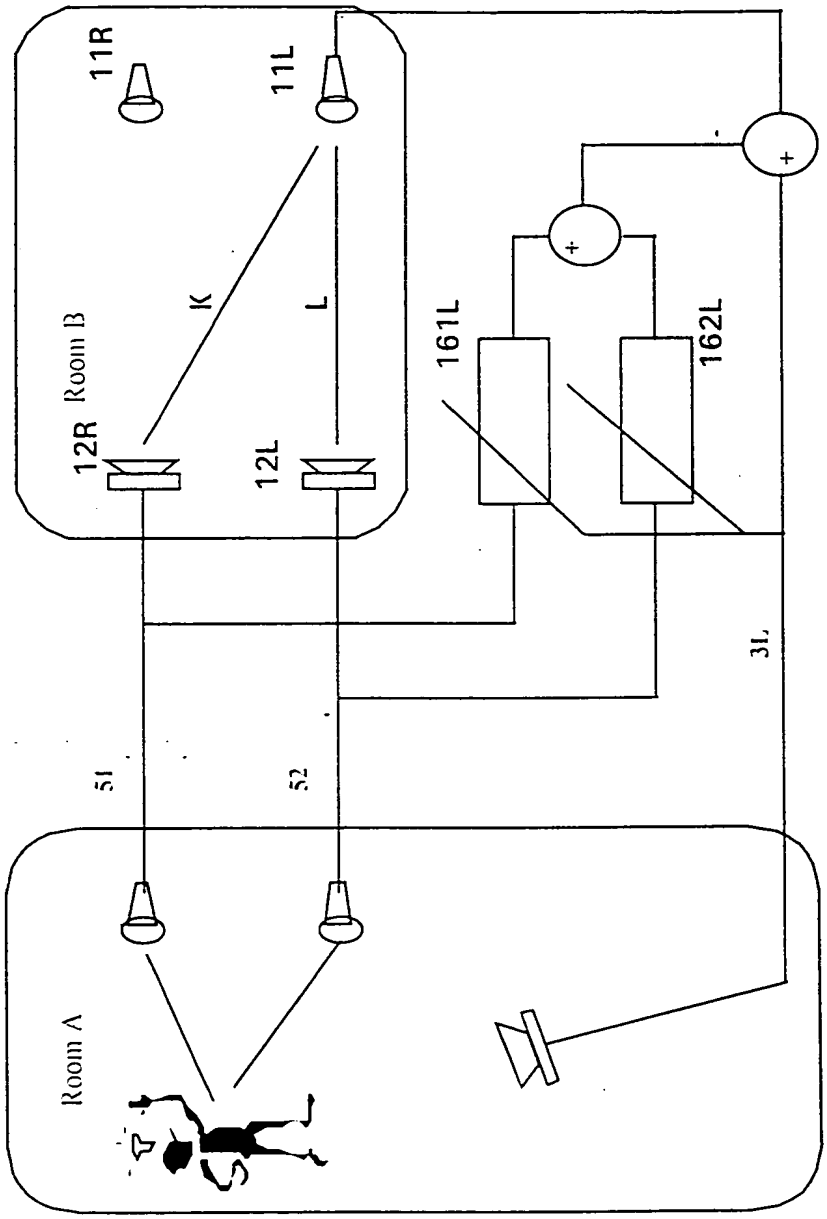

Figure 10

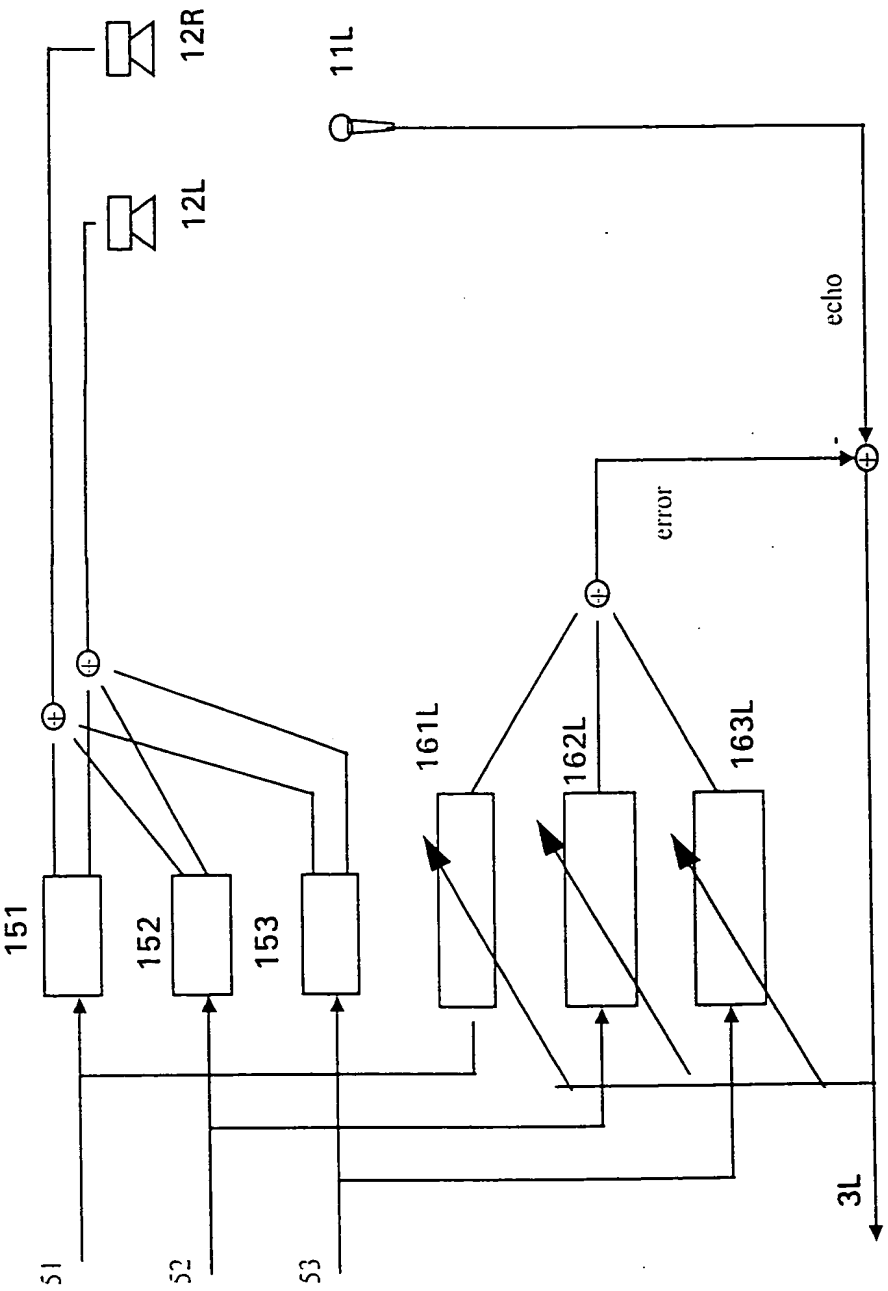

Figure 11

Europäisches Patentamt | European Patent Office | Office européen des brevets

6099/1061

5

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

98302763.2

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

H. Raaphorst

[redacted]AG, DEN
[redacted]GUE, 24/04/99
[redacted]LE

[redacted]EPO/[redacted] Form 1014 - 02.91

Europäisches Patentamt
European Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 98302763.2
Demande n°:

Anmeldetag:
Date of filing: 08/04/98
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
BRITISH TELECOMMUNICATIONS public limited company
London EC1A 7AJ
UNITED KINGDOM

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Teleconferencing system

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
H04M3/36, H04M3/56

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

TELECONFERENCING SYSTEM

This invention relates to audio teleconferencing systems. These are systems in which three or more participants, each having a telephone connection, can participate in a multi-way discussion. The essential part of a teleconference system is called the conference "bridge", and is where the audio signals from all the participants are combined. Conference bridges presently function by receiving audio from each of the participants, appropriately mixing the audio signals, and then distributing the mixed signal to each of the participants. All signal processing is concentrated in the bridge, and the result is monaural (that is, there is a single sound channel). This arrangement is shown in Figure 1, which will be described in detail later. The principal drawback with this system is that the audio quality is monophonic, generally poor, it is very difficult to determine which participants are speaking at any one time, especially when the number of participants is large.

According to the invention, there is provided a teleconferencing system comprising a conference bridge having a multichannel connection to each customer equipment, the customer equipment having means to separately process each channel to provide an output, preferably spatialised, representing each of the other participants. Preferably the conference bridge comprises a concentrator, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels. This reduces the capacity required by the multichannel connection. The control information identifying the active channels may be carried in a separate control channel, or as an overhead on the active subset of channels. In a preferred arrangement the channel representing a given participant is excluded from the output provided to that participant. This may be achieved by excluding that channel from the processing in the customer equipment, but is preferably achieved by excluding it from the multichannel transmission from the bridge to that participant, thereby reducing further the capacity required by the multichannel connection.

Exemplary embodiments of the invention will now be described, by way of example, with reference to the drawings, in which:

Figure 1 illustrates a conventional teleconference system;

Figure 2 illustrates a spatial audio teleconference system according to one embodiment of the invention;

Figure 3 illustrates a N-channel speech decoder used in the embodiment of Figure 2;

Figure 4 illustrates a N-Channel audio spatialiser used in the embodiment of Figure 2;

Figure 5 illustrates a second embodiment of the invention;

Figure 6 illustrates how the invention may be used with conventional PSTN channels;

Figure 7 illustrates a variant of the invention for use with a video conference system;

Figure 8 illustrates a voice switched concentrator which may be used in the embodiments of the invention.

Figures 9, 10, and 11 illustrate various echo cancellation techniques.

In the conventional system illustrated in Figure 1 the conference bridge located in the exchange equipment 100 receives signals from the various customer equipments 10, (20, 30 not shown) in response to sounds detected by respective microphones 11, 21, 31 etc. These signals are transmitted over the telephone network (1), to the exchange 100 at which the bridge is established. Generally the signals will travel by way of a local exchange (not shown) in which the analogue signals are converted to digital form, usually employing linear companding such as "A law" (as used for example in Europe) or "mu-Law" (as used for example in the United States of America) for onward transmission to the bridge exchange 100. On arrival at the bridge exchange 100, the bridge passes each incoming signal 11, 21, 31 through a respective digital converter 111, 112, 113 to convert them from A Law to linear digital signals, and then passes the linear signals to a digital combiner 120 to generate a combined signal. This combined signal is re-converted to A law in a further digital converter 110, and the resulting signal transmitted over the telephone network (2) to each customer equipment 10, (20, 30) for conversion to sound in respective loudspeakers 12, 22, 32 etc. In this way the exchange equipment 100 acts as a "bridge" to allow one or more customer equipments 32 to connect into a simple two-way connection between customer equipments 10, 20.

The systems illustrated in Figures 2 to 8 replace the conventional conference bridge system of Figure 1 with a multicast system in which several channels can be transmitted to each participant, using a multi-channel link comprising an uplink 3, and a downlink comprising a control channel 4 and a digital audio downlink 5 comprising several channels 51, 52. Participants with suitable equipment can then process these channels 51, 52 in various ways as will be described.

The transmission medium used for the uplink 3 and downlink 4,5 can be any suitable medium. ISDN (Integrated Services Data Network) technology or LAN (Local Area Network) - respectively public and private data networks - are the favoured transmission options since they provide adequate data rate and low latency - delays due to coding and transmission buffering. However, they are expensive and so far have a low penetration in the market place. Internet Protocol techniques are more widely used, but have poor latency and unreliable data rates. Being packet systems, they are less suited to voice applications. It is also possible to use the conventional PSTN (public switch telephone system) with a speech band modem. The latest internet type modems provide up to 56kbit/s downstream (links 4,5: digital network down to the customer via local loop), and up to 28.8kbit/s upstream (link 3). They are low cost and are commonly bundled into PC packages. Ideally a system should be able to work with all of the above, and with standard analogue PSTN available as a backup.

The signal mixing can take place either in the user's terminal equipment, or in a centralised processing platform as shown in Figure 2. In Figure 2 the customer equipment 10 contains a microphone 11 and loudspeaker system 12 as before. However, the loudspeaker system 12 is a spatialised system - that is, it has two or more channels to allow sounds to appear to emanate from different directions. This may take the form of stereophonic headphones, or a more complex system such as disclosed in United States Patents 5533129 (Gefvert), 5307415 (Fosgate), article *"Spatial Sound for Telepresence"* by M.Hollier, D. Burraston, and A. Rimell in the *British Telecom Technology Journal,* October 1997 or the applicant's own pending European Patent Application 97304218.7 filed on 17th June 1997.

The output from the microphone 11 is encoded by an encoder 13 forming part of the customer equipment 10, and transmitted over the uplink 3 to the

exchange equipment 100. Here it is combined with the other input channels 21, 31 from the other participants into a concentrator 230 which combines the various inputs into an audio signal having a smaller number of channels 51, 52. These channels are transmitted over multiple-channel digital audio links 5 to the customer equipments 10, (20, 30) where they are first decoded by respective decoders 14, 24, 34 and provided to a spatialiser 15 for controlling the mixing of the channels to the generate a spatialised signal in the spaeker equipment 12.

The concentrator 230 selects from the input channels 11, 21, 31 those carrying useful information - typically those carrying speech, and passes only these over the return link 5. This reduces the amount of information to be carried. A control channel 4 carries data identifying which channels were selected. The spatialiser 15 uses data from the control channel to identify which of the original sound channels 11, 21, 31 it is receiving, and on which of the "N" channels 51, 52 in the audio link each original channel is present, and constructs a spatialised signal using that information. The spatialised signal can be tailored to the individual customer, for example the number of talkers in the spatialised system, the customer's preferences as to where in the spatialised system each participant is to appear to be located, and which channels to include; for example the original talker or a simultaneous translation. In particular, the user may exclude the channel representing his own input 11.

Transmission efficiency is achieved because only the active subset N of the total number of channels M are transmitted at any one time. The subset is chosen using a voice controlled dynamic channel allocation algorithm in the N:M concentrator 230. A possible implementation of this is shown in Figure 8. Each input channel 11,21,31 is monitored by a respective analyser 231, 232, 233. As shown for analyser 231, the signal is subjected to a speech detection and analysis process 231b. This detects whether speech is present on the respective input 11, and gives a confidence value, indicative of how likely the signal contains speech. This ensures that low-level background speech is given a lower weight than speech clearly addressed to the microphones 11, 21, 31 etc. A value is also given for level, to ensure speech directed to the microphone is preferred over background noise, and the level information can be passed to the spatialistion system to select a coding algorithm appropriate to the information in the speech. In order to detect

and process the speech in the signals they first need to be decoded in a decoder 231a (this may be dispensed with if the speech detection system 231b can operate with digitally encoded signals).

A voting algorithm 234 then selects which of the inputs 11, 21, 31 have the clearest speech signals and controls a switch to direct each of those input channels 11, 21, 31 which have been selected to a respective one of the output channels 51, 52. Similar algorithms are used in Digital Circuit Multiplication Equipment (DCME) systems in international telephony circuits. Data relating the audio channels' content to the conference participants, and therefore the correspondence between the input channels 11, 21, 31 and output channels 51, 52 is transmitted over the control channel 4. Alternatively, this data can be embedded in the encoded audio data.

When there are fewer talkers identified than there are available output channels 51, 52, signal quality can be improved by using a less compressed digitisation scheme for those input channels selected, thereby using more than one output channel 51, 52 for each input channel selected. Telephone quality speech may be achieved at 8kbits/s, allowing eight talkers to be accommodated if the system has a 64kbit/sec capability. Should fewer talkers be detected, the 64kbit/s capability may be used instead to provide four 16 kbit/s audio channels, capable of carrying 'good' quality speech, or a mixture of channels at different bit rates, to allow the coding rates to be selected according to the initial signal quality, or so that the main talker could be passed at higher quality than the other talkers. Layered coding schemes can be used to allow graceful switching between data rates.

The N-channel de-multiplexer and speech decoder 14 is shown in Figure 3. This receives the channels 51, 52, 53 etc carried in the audio downlink 5 and separates them in a demultiplexer 140. Each channel 51, 52, etc is then separately decoded in a respective decoder 141, 142, 143, etc for processing by the spatialiser 15. The decoders 141, 142, etc may operate according to different processes according to the individual coding algorithms used, under the control of the control signals carried in the control channel 4.

A composite signal consisting of the summation of all input signals could also be transmitted. Such a signal could be used by users having monaural

receiving equipment, and may also be used by the spatialiser to generate an ambient background signal. Alternatively the spatialiser 15 may replace any of the channels 11, 21, 31 not selected by the concentrator 230, and therefore not represented in the N channel link 5, by "comfort noise"; that is low-level white noise, to avoid the aural impression of a void which would be occasioned by a complete absence of signal.

The customer equipment 10 can be implemented using a desktop PC. Readily available PC sound card technology can provide the audio interface and processing needed for the simpler spatialising schemes. The more advanced sound cards with built in DSP technology could be used for more advanced schemes. The spatialiser 15 can use any one of a number of established techniques for creating an artificial audio environment as detailed in the Hollier et al article referred to above. Spatialisation techniques may be summarised as follows. The simplest technique is "panning", where each signal is replayed with appropriate weighting via two or more loudspeakers such that it is perceived as emanating from the required direction. This is easy to implement, robust and may also be used with headphones.

"Ambisonic" systems are more complicated and employ a technique known as wavefront reconstruction to provide a realistic spatial audio percept. They can create very good spatial sound, but only for a very small listening area and are thus only appropriate for single listeners. For headphone listening, "binaural" techniques can be used to provide very good spatialisaton. These use head-related transfer function (HRTF) filter pairs to recreate the soundfield that would have been present at the entrance to the ear canal for a sound emanating from any position in 3D space. This can give very good spatialisation and may be extended for use with loudspeakers, when it is known as "transaural". As with ambisonic systems, the correct listening position is very small. Any of these spatialisation techniques may be used with the present invention.

The output of several spatialisers may be combined as shown in Figure 4, which shows a spatialiser group for a stereophonic output having left and right channels 12L, 12R. Each channel 51, 52, 53 is fed to a respective spatialiser 151, 152, 153 which, under the control of a coefficient selector 150 control by the signals in the control channel 4, transmits an output 151L, 151R etc to each of a

series of combiners 15L, 15R. The processing used to create the outputs 151L, 151R etc is operated under the control of the signal 4 such that each channel appears as a virtual sound source, having its own location in the space around the listener.

The positions of virtual sources in three dimensional space could be determined automatically, or by manual control, with the user selecting the preferred positioning for each virtual sound source. For a video conference the positioning can be set to correspond with the appropriate video picture window. The video images may be sent by other means, or may be static images retrieved from a local storage by the individual user.

If the spatialised sound is relayed via loudspeakers 12, rather than headphones, it will be necessary to prevent signals from the loudspeakers 12 being picked up by the microphone 11, re-transmitted and being heard as an echo at the distant sites 20, 30 etc. A technique for achieving this will be described later, with reference to Figure 11.

Figure 5 shows an alternative arrangement to that of Figure 4, in which the spatialisation is computed in the conference 'bridge'. Each conference participant gets the same spatialised signals, thus simplifying the customer equipment. Figure 5 is similar in general arrangement to Figure 2, except that the decoder 14 and spatialiser 15 are part of the exchange equipment 200. The output from the spatialiser 15 is passed to an encoder 18 which transmits the required number of audio channels (e.g. two for a stereo system) to each customer 10, 20, 30. This requires the number of channels in the downlink 5 to be equal to the number of audio channels in the spatialisation systems' outputs, instead of the number selected by the concentrator (plus the control channel 4) as in the embodiment of Figure 2. It also simplifies the customer equipment 10. However, this arrangement requires all customer installations 10, 20, 30 to have similar spatialisation systems, and in particular the same number of audio channels. It would also be more difficult to remove a talker's own voice from the signal he receives. Echo control would also be more complicated, and channel coding may degrade the spatialisation.

Conventional analogue connections could be included in the conference by providing each analogue connection 43, 45 to the 'bridge' 200 with an encoder

42, as shown in Figure 6, to provide an input 41 to the concentrator. The output 5 of the concentrator 230 is also decoded and combined in a unit 44 to provide a monaural conference signal 45 to the analogue user 40.

In the embodiments of Figures 2 to 4 and Figure 5, if loudspeakers are used there is a need to control acoustic feedback ("echo") between the loudspeaker 12 and the microphone 11, which will result in signals being retransmitted back into the system. This will result in each user hearing one or more delayed versions of each signal (including their own transmissions) arriving from the other users. For a monophonic system echo control is usually done using an echo canceller as shown in Figure 9. The echo signal, represented by D is caused by the acoustic path J between the loudspeaker 12 and microphone 11 of equipment 10 in room B. The cancellation is achieved in an echo control unit 16 by using an adaptive filter to create a synthetic model of the signal path such that the echo may be removed by subtraction. The signal E, returned to equipment 20 in room A, is now free of echoes, containing only sounds that originated in Room B. The optimum modelling of the acoustic path J is usually achieved by the adaptive filter in a manner such that some appropriate function of the signal E is driven towards zero. Echo control using adaptive filters in this manner is well known.

Multi-channel echo cancellation, as shown in Figure 10 for two channels, is more complex since there are two input channels 51, 52 and therefore two loudspeakers 12L, 12R. It is therefore necessary to model two echo paths K and L for each of the two return channels 3L, 3R. (The process is only shown for return channel 3L, using microphone 11L). Correct echo cancellation is only achieved if adaptive filters 161L, 162L model the signal paths K and L respectively. (Two further filters 161R, 162R are required for the other return channel 3R) However, it is not possible to find a correct model for each path K, L independently without some difficult and expensive signal processing as described in *"A better understanding and an improved solution to the specific problems of stereophonic echo cancellation"* (IEEE Transactions on speech and Audio processing, Vol 6, no 2 March 1998. Authors: J Benesty, D R Morgan and M M Sondhi).

The system described above with reference to Figure 4 employs linear artificial spatialisation techniques. Figure 11 shows how this, and the fact that the

echo from each loudspeaker 12L, 12R combines linearly at each microphone 11L, (11R, not shown), allows echo cancellation to be provided for each output channel 3L, (3R) by having a separate adaptive filter 161L, 162L, 163L, (161R, 162R, 163R) on each input channel 51, 52, 53. Thus the adaptive filter 161L will model the combination of the spatialiser 151 for the channel 51, and the echo path between the loudspeakers 12L and 12R and the microphone 11L.

The invention could be applied to a conference situation in which there are several participants at each location, such as the video conference shown in Figure 7. Close microphones 11a, 11b, 11c, for example of the "tie-clip" type, are used to pick up the sound from each individual talker, and a talker location system 60 is used to keep track of their spatial position. The talker location system 60 may comprise a system of microphones which can identify the positions of sound sources. Relating the position of a sound source to that of the tie clip microphone 11 currently in use makes it possible to learn the position of each talker by audio means alone. Alternatively, the system may detect the position of each user by means such as optical recognition of a badge carried by each user. In either case, the position data is passed to the far end (Room B), where correct spatialisation is reconstructed, for output by loudspeakers 12L, 12M, 12R etc. This would achieve a true spatial conference and overcome the associated echo control problems, since the "tie clip" microphones 31a, 31b, 31c have a limited range and will not detect the outputs from the loudspeakers 32 in the same room.

3

CLAIMS

1. A teleconferencing system comprising a conference bridge having a multichannel connection to each customer equipment, at least one customer equipment having means to separately process each channel to provide a plurality of outputs, each output representing one of the other participants.

2. A system according to claim 1, wherein the customer equipment has means to combine the outputs representing each participant to provide a spatialised output in which each participant is represented by a virtual sound source.

3. A system according to claim 1 or 2, wherein, the conference bridge comprises a concentrator, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

4. A system according to any preceding claim, wherein the channel representing a given participant is excluded from the output provided to that participant.

5. A system according to claim 4, comprising means in the customer equipment for excluding the said channel from the processing.

6. A system according to claim 4, comprising means for excluding the said channel from the multichannel transmission from the bridge to the respective participant.

7. A system according to any preceding claim, the customer equipment having echo cancellation means comprising means for detecting correlations between the output signal from the customer equipment and the signals carried on individual input channels to the customer equipment representative of other users,

such correlations being indicative of acoustic feedback at the customer equipment, and means for cancelling such feedback signals in the output signal.

8. A system according to claim 7, wherein the customer equipment comprises, for each channel of the output signal, a plurality of adaptive filters, each adaptive filter being arranged to model the echo path between a respective input channel and the respective output channel, and
for each output channel there being provided a combiner for adding the outputs of the respective plurality of adaptive filters to generate an echo cancellation signal for the respective output channel.

9. A method of providing teleconferencing services to a plurality of customer equipments, in which a multichannel connection is provided from a conference bridge to each customer equipment, in which at least one customer equipment processes each channel separately to provide a plurality of outputs, each representing one of the other participants.

10. A method according to claim 9, wherein the conference bridge identifies the currently active input channels and transmits only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

ABSTRACT

TELECONFERENCING SYSTEM

A teleconferencing system comprises a conference bridge 100 having a multichannel connection 5 to each customer equipment 10. The customer equipment 10 has means 14 to separately process each channel 51, 52, 53 to provide one output representing each of the other participants. These outputs can be combined in a spatialiser 15 to provide a spatialised output 12 in which each participant is represented by a virtual sound source. The conference bridge 100 comprises a concentrator 230, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels.

Figure (2)

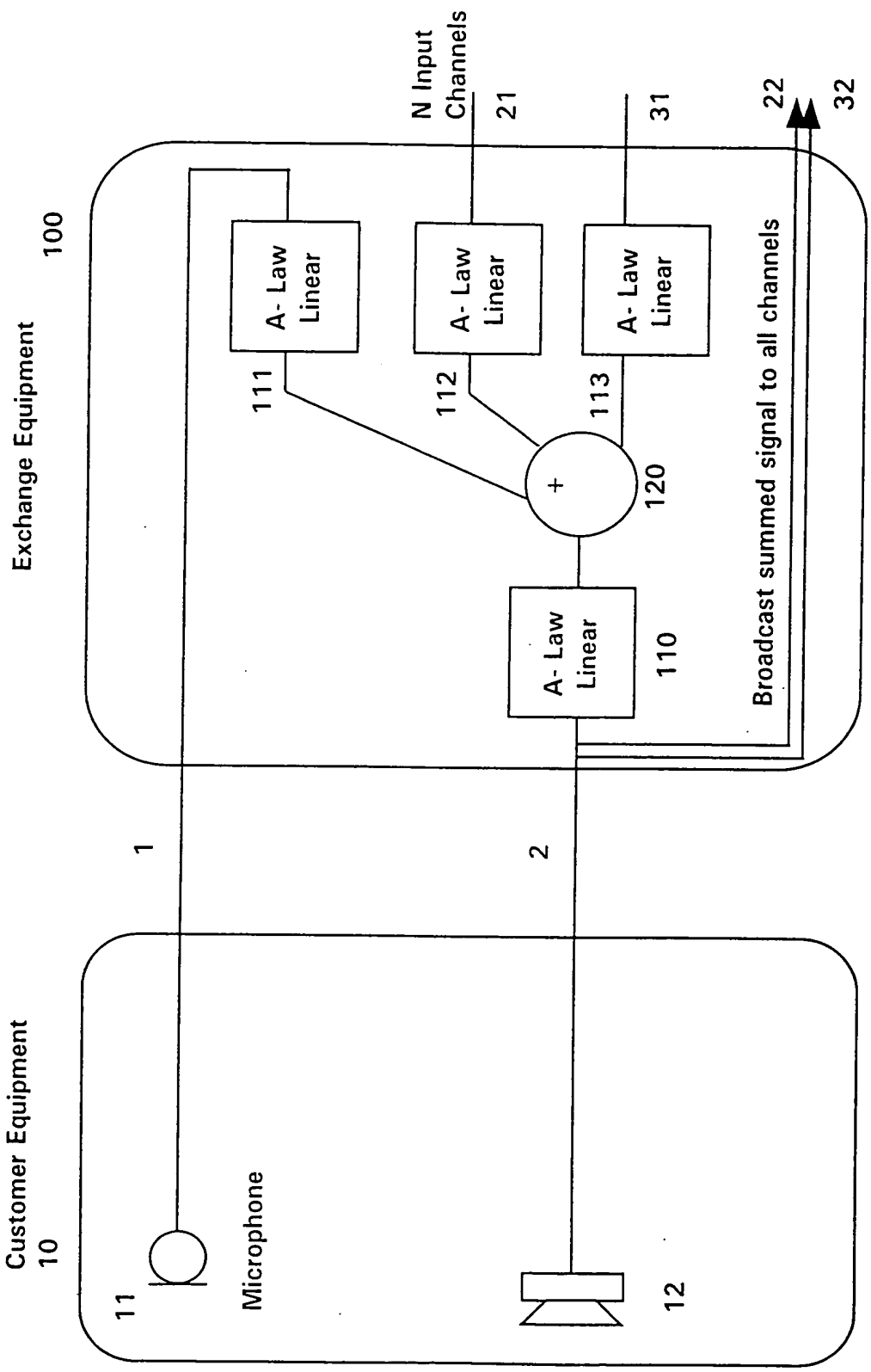
Customer Equipment
10
Microphone
11
12
1
2
Exchange Equipment
100
N Input
Channels
21
31
22
32
A- Law
Linear
111
112
113
110
120
+
Broadcast summed signal to all channels

Figure 1

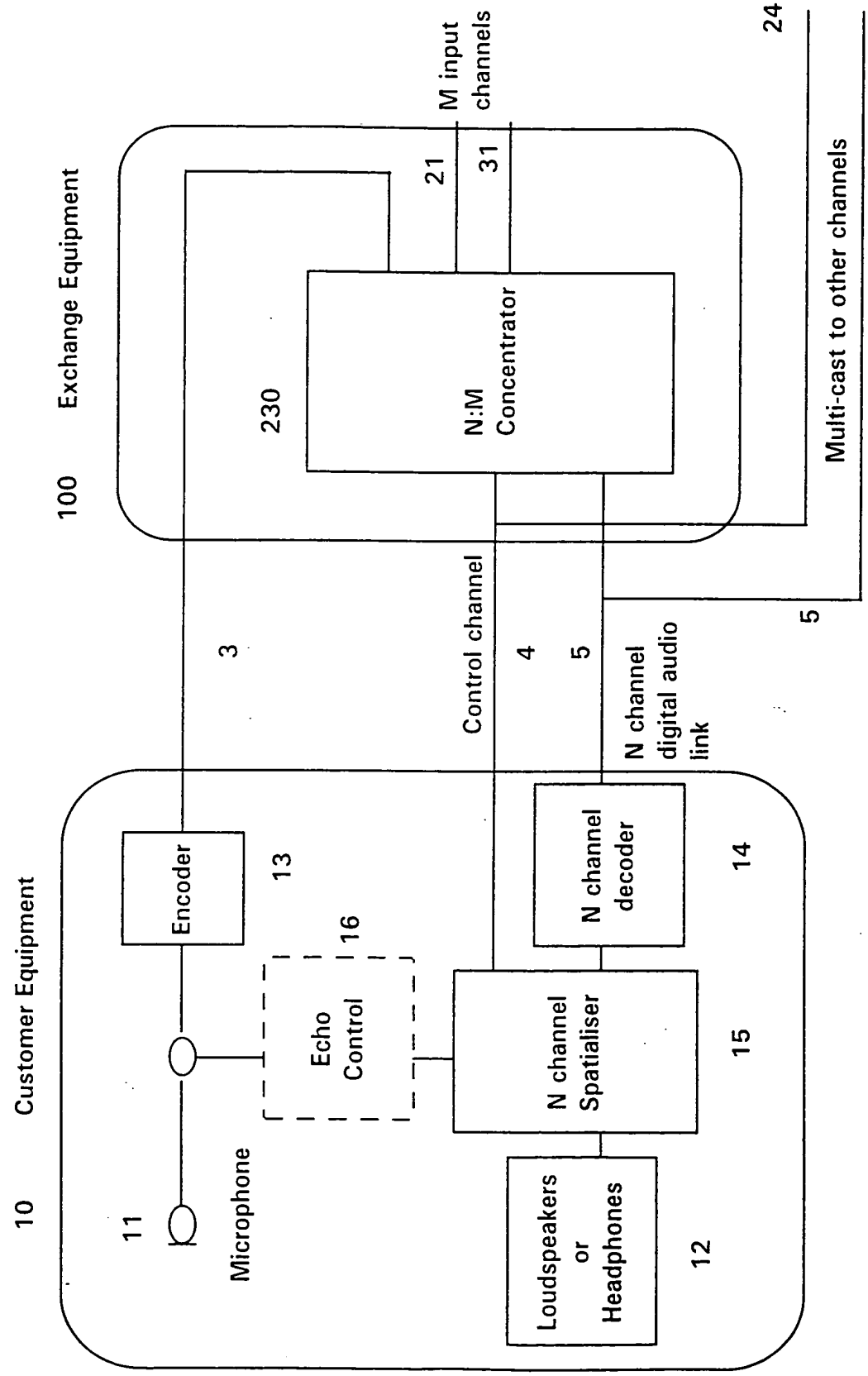
10
Customer Equipment
11
Microphone
Encoder
13
Echo Control
16
N channel Spatialiser
15
Loudspeakers or Headphones
12
N channel decoder
14
3
Control channel
4
5
N channel digital audio link
100
Exchange Equipment
230
N:M Concentrator
21
31
M input channels
24
5
Multi-cast to other channels
34

Figure 2

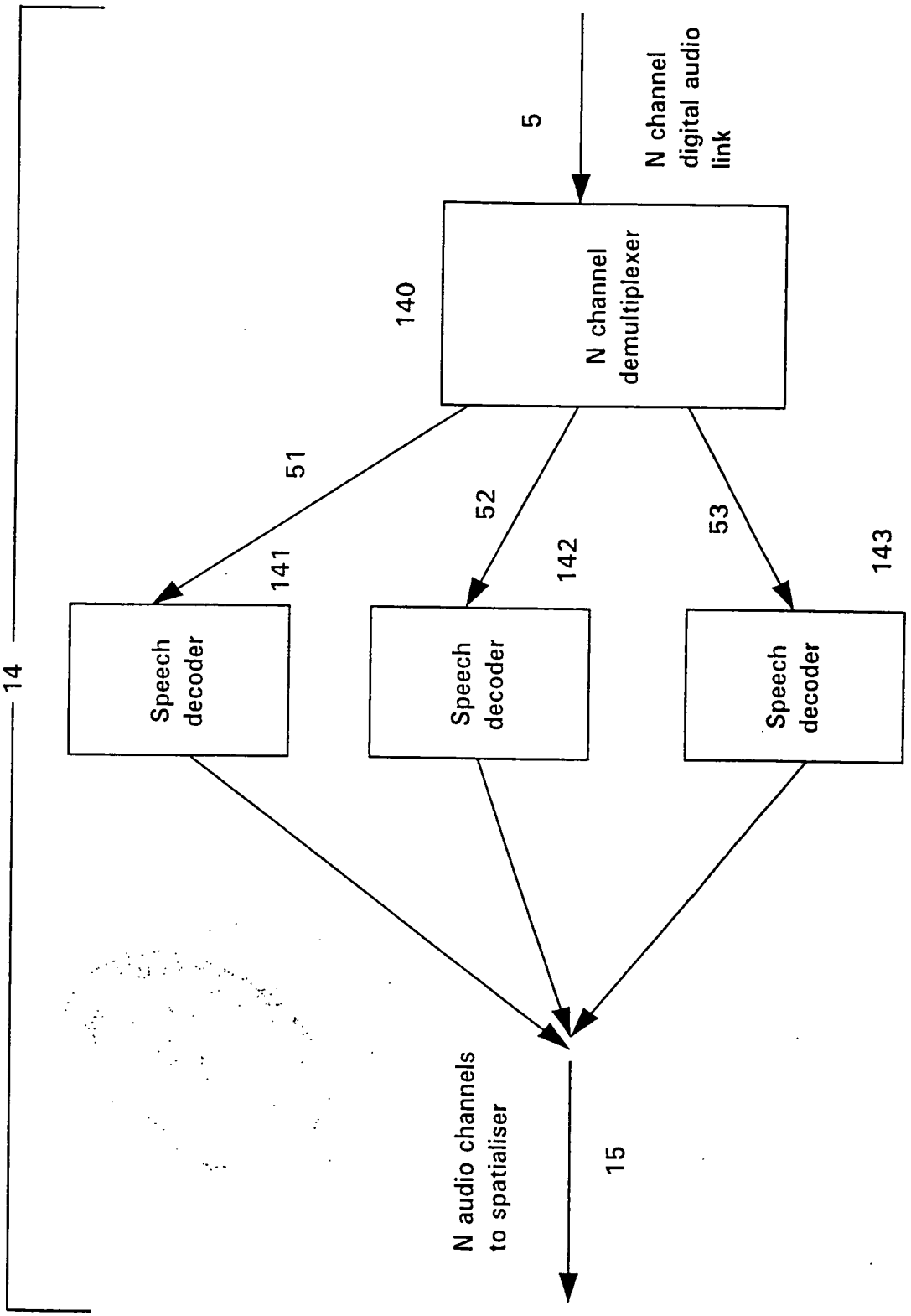
14
5
N channel digital audio link
140
N channel demultiplexer
51
52
53
141
142
143
Speech decoder
Speech decoder
Speech decoder
N audio channels to spatialiser
15

Figure 3

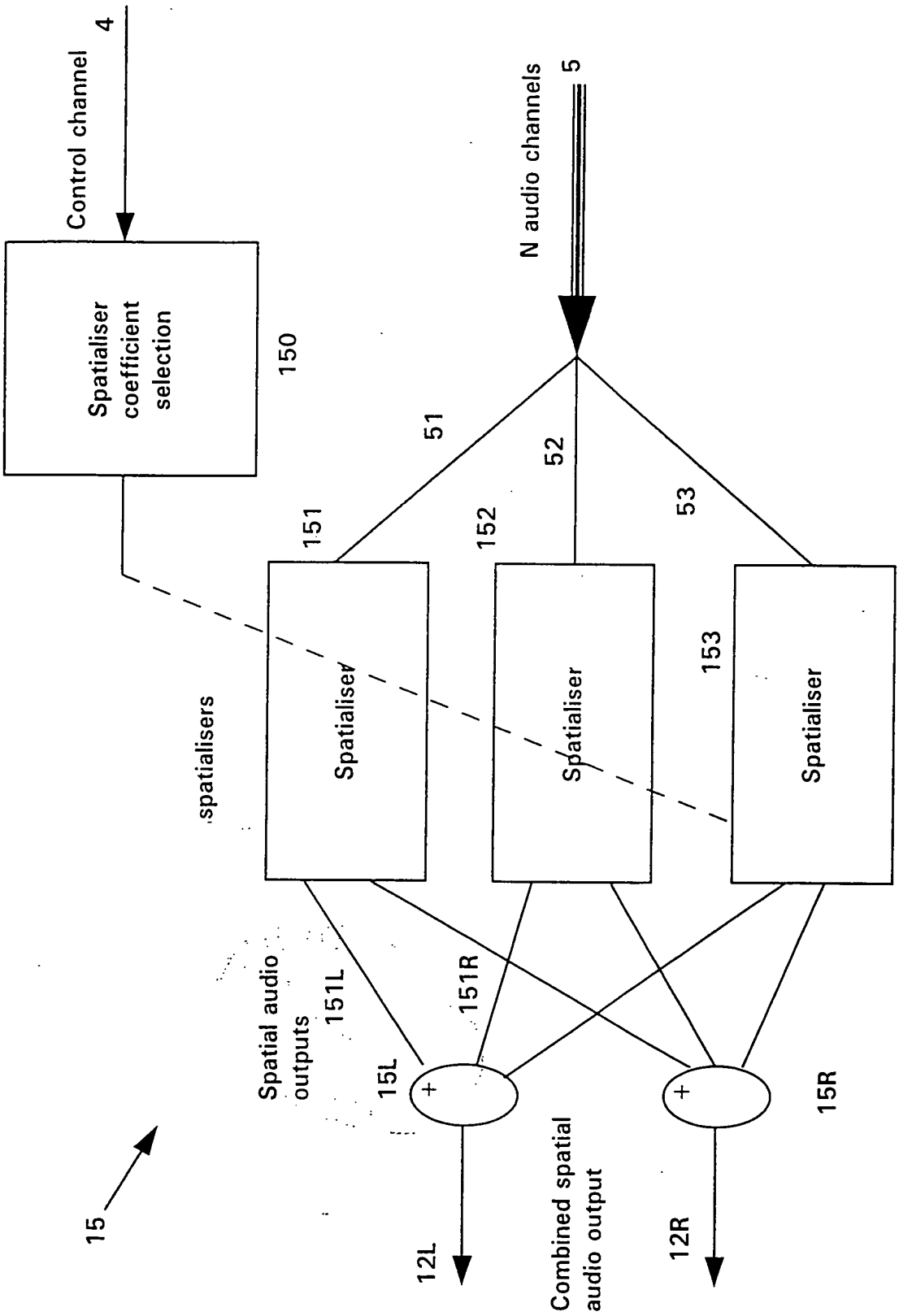
Control channel
4
Spatialiser coefficient selection
150
N audio channels
5
51
52
53
151
152
153
spatialisers
Spatialiser
Spatialiser
Spatialiser
Spatial audio outputs
151L
151R
15L
15R
+
+
12L
12R
Combined spatial audio output
15

Figure 4

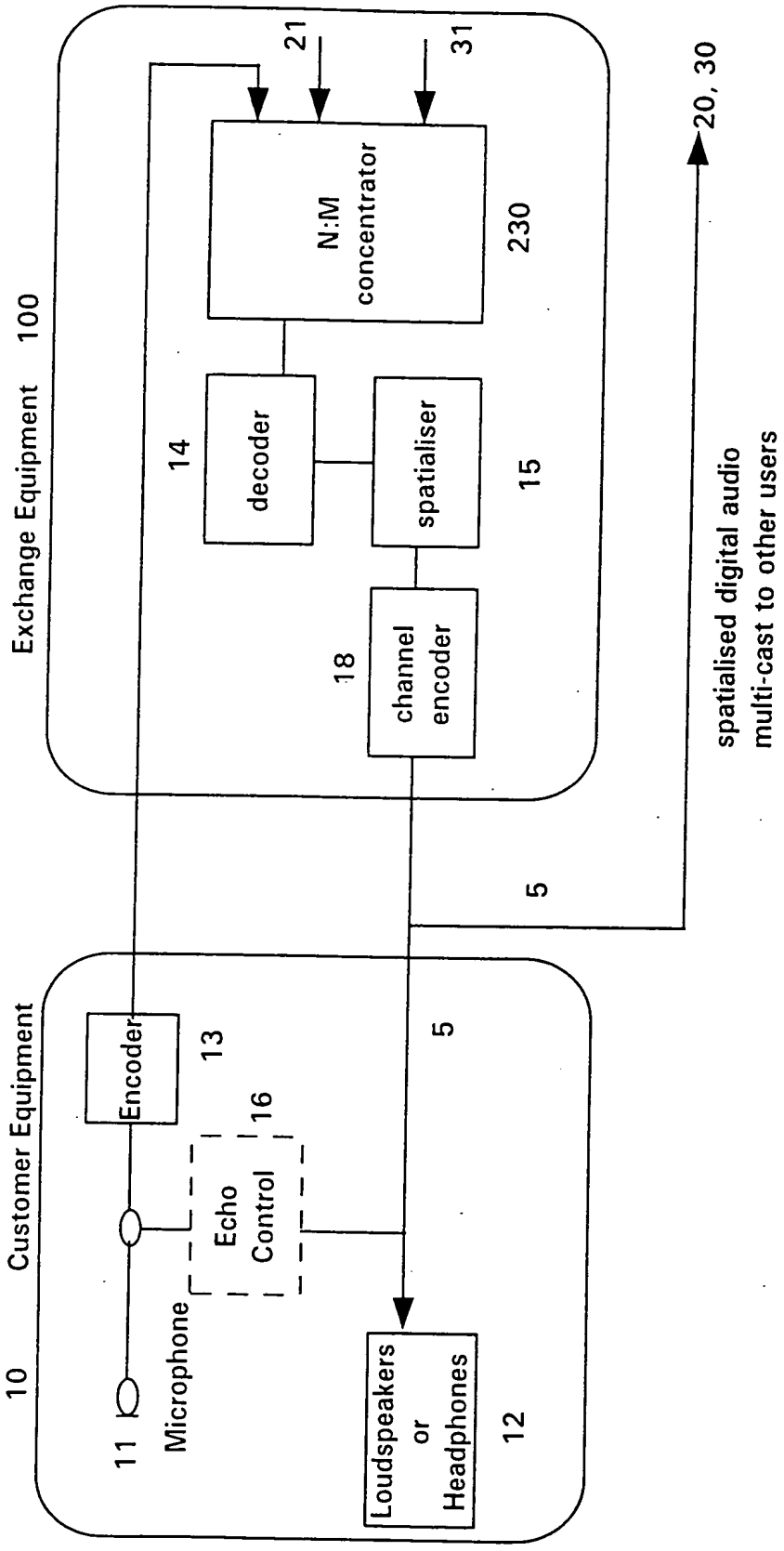
Exchange Equipment 100
21
31
N:M concentrator
230
decoder
14
spatialiser
15
channel encoder
18
20, 30
spatialised digital audio multi-cast to other users
5
Customer Equipment
10
Encoder
13
16
Echo Control
Microphone
11
5
Loudspeakers or Headphones
12

Figure 5

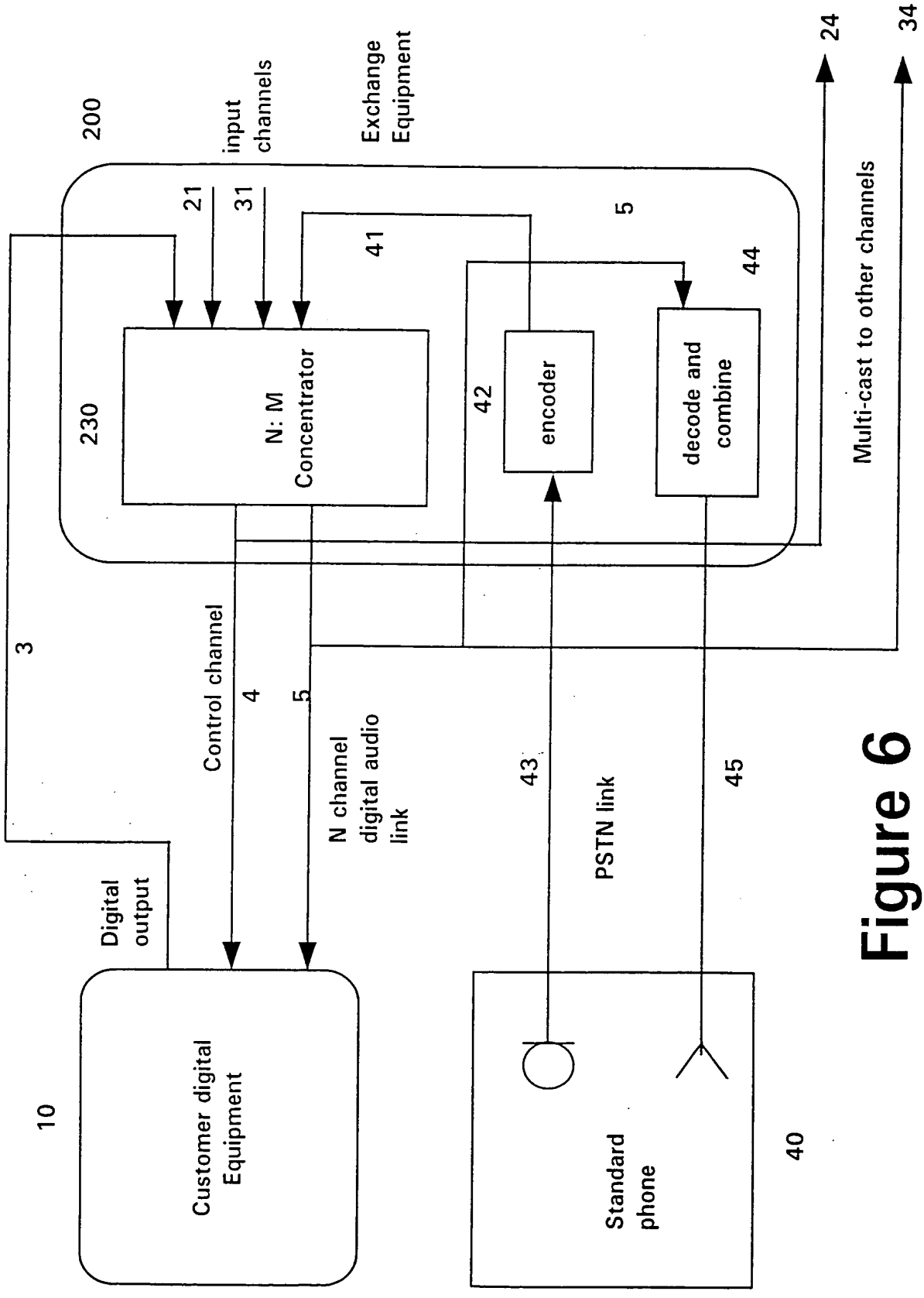
200
Exchange Equipment
input channels
21
31
41
5
44
24
34
230
N: M Concentrator
42
encoder
decode and combine
Multi-cast to other channels
3
Control channel
4
5
N channel digital audio link
43
PSTN link
45
Digital output
10
Customer digital Equipment
Standard phone
40

Figure 6

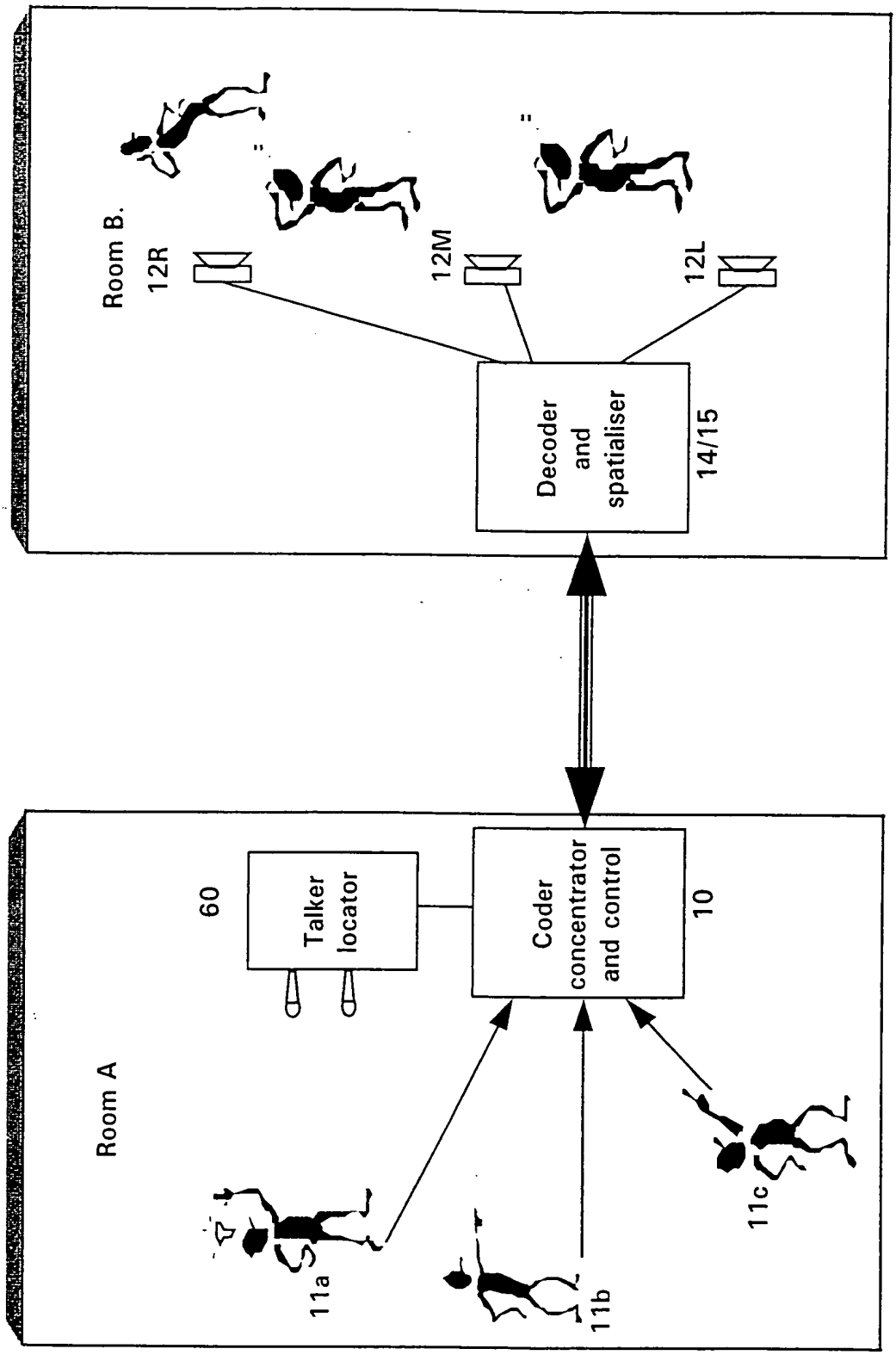
Room B.
12R
12M
12L
Decoder and spatialiser
14/15
Room A
60
Talker locator
Coder concentrator and control
10
11a
11b
11c

Figure 7

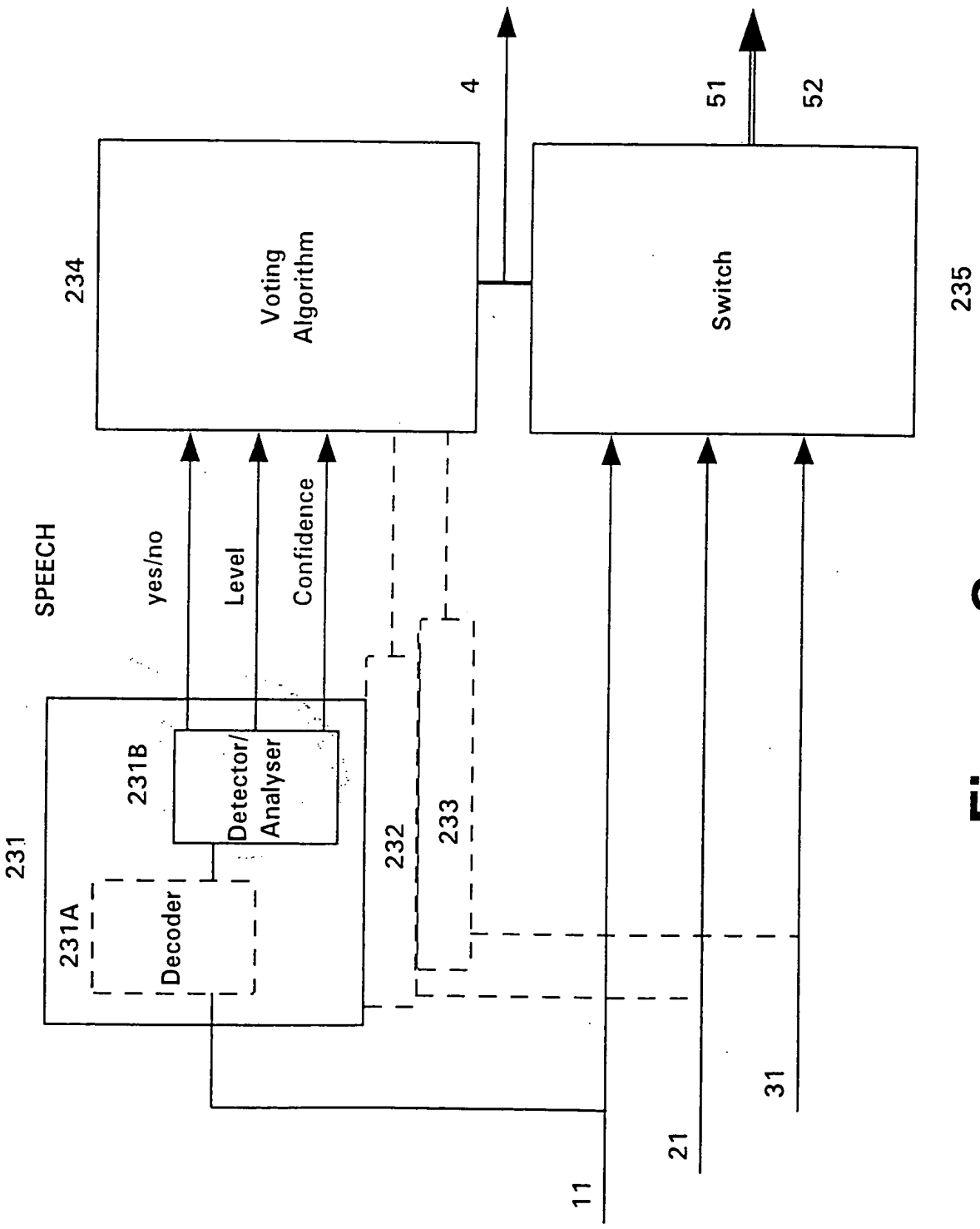
SPEECH
yes/no
Level
Confidence
231
231A
Decoder
231B
Detector/
Analyser
232
233
234
Voting
Algorithm
4
51
52
Switch
235
11
21
31

Figure 8

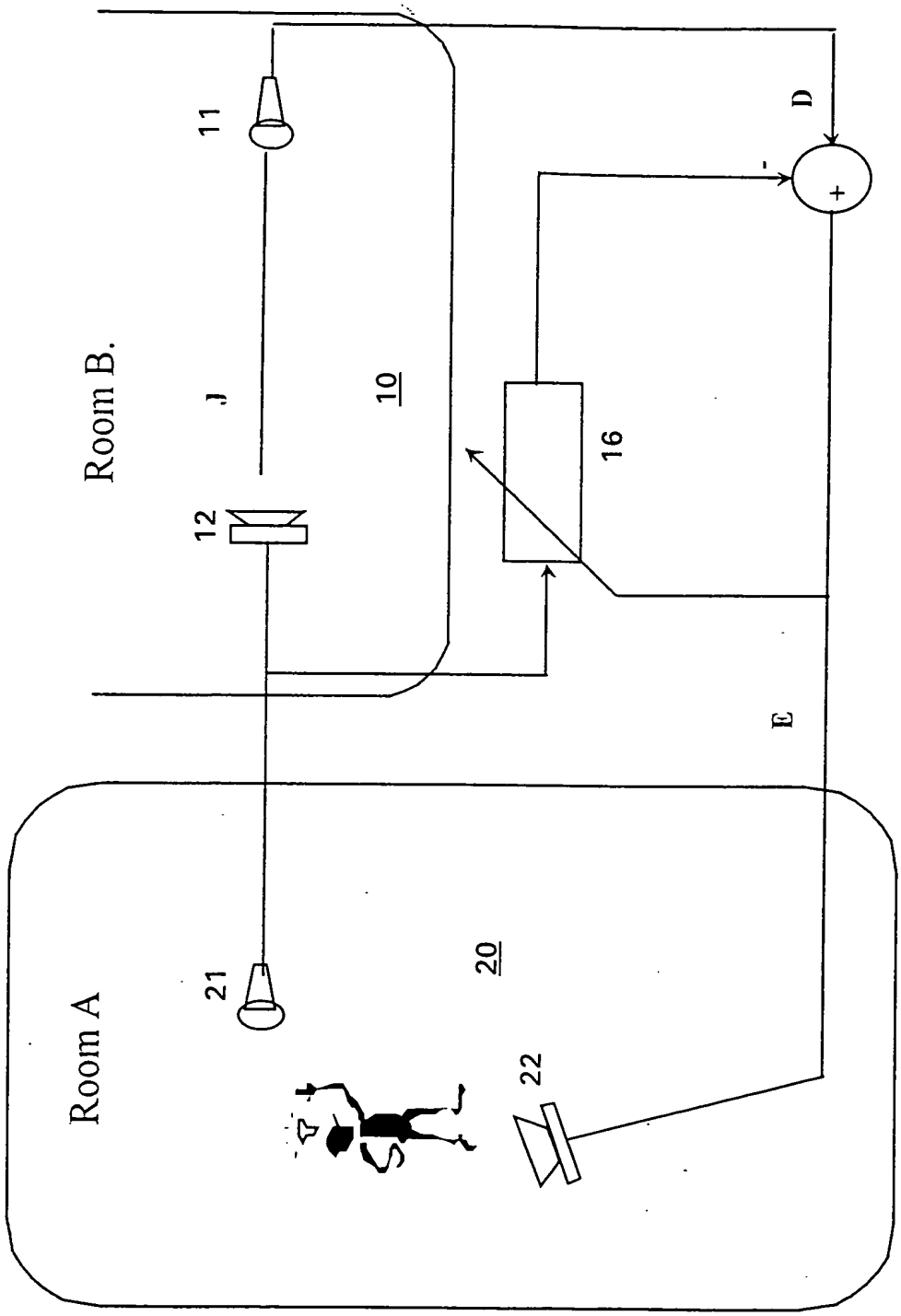

Figure 9

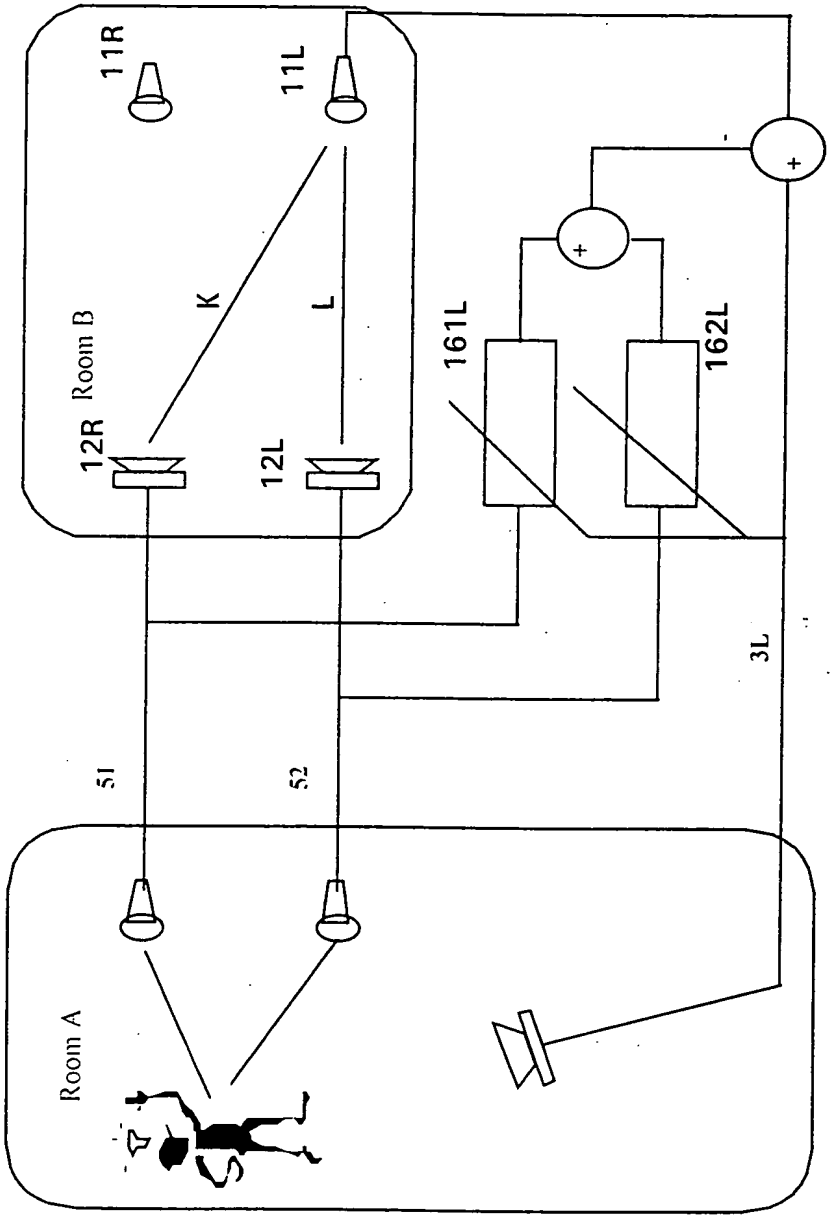

Figure 10

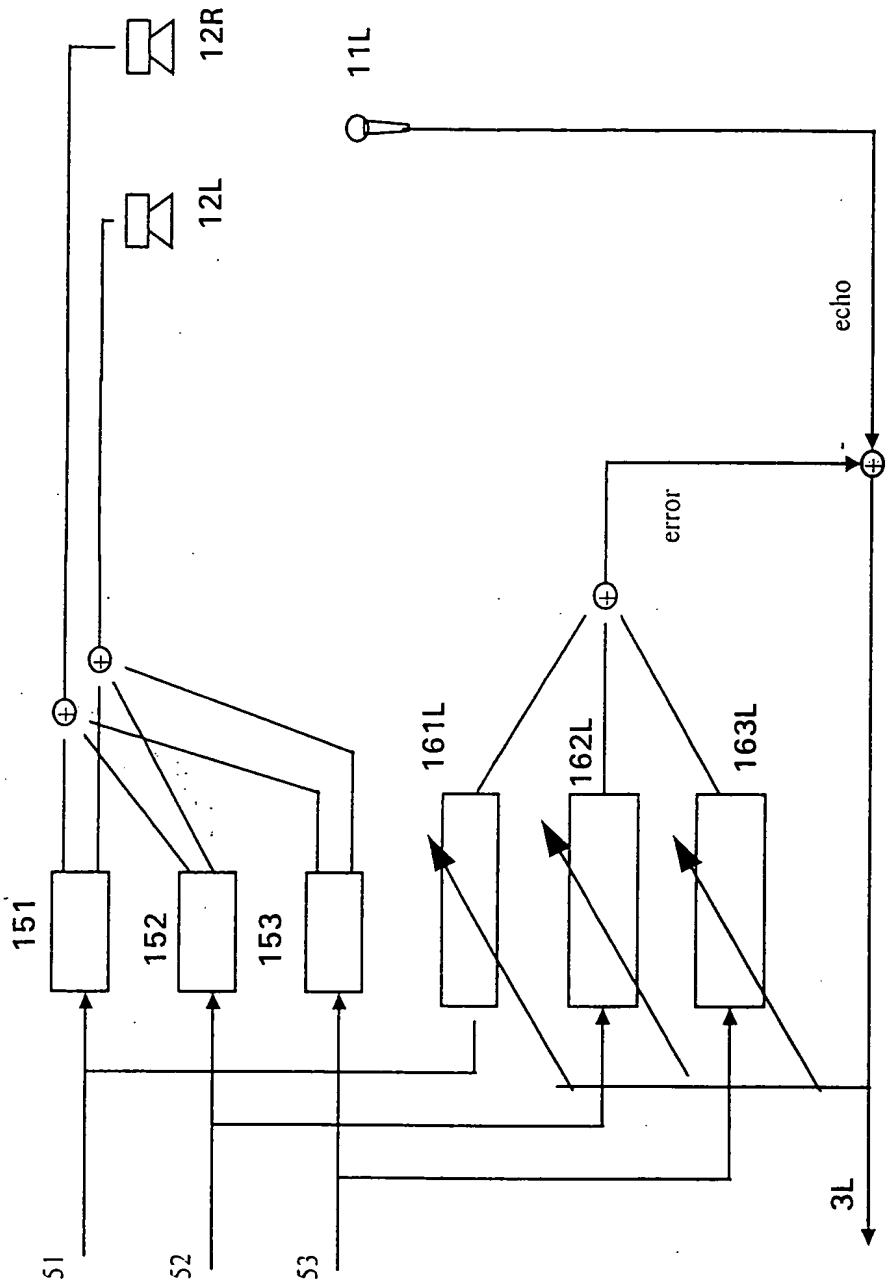

Figure 11